The Dynamics of Internal Migration: A New Fact and its Implications*

Greg Howard† Hansen Shao‡

February 20, 2025

Abstract

We show a new fact: the t-year interstate migration rate is close to proportional to the square root of t. This is a puzzle for simple moving cost models which predict an approximately linear relationship. We propose a model based on persistent and spatially-correlated unobservable match-specific characteristics that matches the square root fact, as well as the frequency and gravity pattern of internal migration. Compared to the standard model, the new model is better at forecasting individuals' locations, has different implications for welfare effects of migration policies, and makes different predictions for long-run population changes in response to shocks.

Keywords: regional evolution, misallocation, gravity equation, labor mobility, moving costs

JEL Codes: R23, R13, J61

*This paper was previously circulated as "Internal Migration and the Microfoundations of Gravity." We would like to thank Javier Quintana, Kyle Mooney, Yao Wang, Vivek Bhattacharya, Gregor Schubert, Andrii Parkhomenko, Eduardo Morales, Marieke Kleemans, and participants at the Illinois Macro Lunch, Illinois Young Applied Faculty Lunch, the European UEA meetings, Midwest Macro, the AREUEA National Meetings, the Stanford Institute for Theoretical Economics Housing & Urban Economics Conference, the Online Spatial and Urban Seminar, the Chicago Fed Urban Conference, the Cleveland Federal Reserve, and the North American UEA meetings for constructive feedback. We also would like to thank Jialan Wang, Julia Fonseca, and Peter Han for creating the Gies Consumer and Small Business Credit Panel and the Gies College of Business for supporting this dataset.

†University of Illinois Urbana-Champaign, glhoward@illinois.edu.

‡University of Illinois Urbana-Champaign.

We document a new empirical regularity: the t-year interstate migration rate, defined as the share of people living in a different state than they did t years ago, scales quite closely with the square root of t. This new fact is a puzzle for the standard moving cost model that is widely used in the literature, which typically implies a linear relationship.

The main contribution of our paper is a simple but novel model that can match the new fact by assuming that individual-location-match-specific utility is correlated across time and space. The model we propose features Spatially and Persistently Auto-Correlated Epsilons, so we call it the "SPACE" model.[1] Unlike the standard model, which understands low migration rates as the result of large moving costs, our model rationalizes low migration rates as a result of the match-specific determinants of location being highly persistent. Our model is able to replicate bilateral one-year migration flows from the data, maintaining the flexibility of the standard models, while also featuring more realistic dynamics.

With the new model in hand, we compare the implications of the new model to the standard moving cost model and find that, for many but not all questions, the models have different answers. These findings reshape our understanding of the causes of low migration, the dynamics of population adjustment, the long-run population elasticities to local changes, and the changes in implied utilities across space in recent years.

The first part of our paper documents the new square root fact using data from the Gies Consumer Credit Panel, a 15-year panel recording the location of approximately 1 percent of all Americans with a credit report every year. While the square root fact is related to some other well-known facts, such that return migration is common (Kennan and Walker, 2011), we show that the $\sqrt{t}$ fact is not a simple result of return migration, but captures richer dynamics. We also show that this fact does not naturally occur in standard moving cost models. In those models, migration is a Markov process, which combined with low rates of migration, implies a linear relationship with t.[2]

We then build a model that can reconcile this new fact while maintaining the tractability and flexibility of the standard model. Building on the model of McFadden (1977), we consider a generalized extreme value discrete choice model to introduce correlation over space and time. The SPACE model leads to closed-form solutions for state populations

[1] ϵ is the common notation for the random component in a random utility model.

[2] With flexible enough moving costs that change over the time a person lives in one spot, a model could match almost any relationship between the t-year migration rate and t. Nonetheless, we still think the square root fact is a puzzle because for two reasons: first, the models that people actually use do not explain it, possibly due to the large state-space of very flexible models; and second, because when a model can match any relationship, there is no reason for it to match this particular relationship. Our model provides a rationale for this particular relationship. We discuss this further in Section 2.1.

and interstate migration. One important result of this model is that the cross-state population elasticities—a key statistic for quantitative spatial modeling—are directly proportional to the bilateral migration rate.

We show a way to calibrate the model that allows for simple formulas for population changes, i.e. "exact hat" algebra, and which also allows computationally-feasible simulations of an individuals' location choices over time.[3]

Next, we compare the implications of the SPACE model to those of the moving cost model. For many questions, we show that choosing how to model migration is not an innocuous choice, but leads to important differences in how economists answer central questions about location choice and migration.

First, we show that the SPACE model is better at predicting future locations of individuals. We compare the forecasting performance of each model using simulated log likelihood, and demonstrate that the SPACE model does better at predicting out-of-sample locations, especially at longer-horizons. This is consistent with the idea that the SPACE model is able to match realistic dynamics of location choice.

Second, the SPACE model and the moving cost model have very different perspectives on why people do not move. Moving cost models estimate large moving costs (e.g. Kennan and Walker, 2011; Giannone, Li, Paixao and Pang, 2020; Zerecero, 2021). In contrast, the SPACE model does not need moving costs at all to rationalize observed levels of migration in the data. To the extent that one thinks of moving costs as a friction that causes misallocation and which can be overcome by policy, those potential welfare gains are not present in the SPACE model.[4]

Third, we turn to macroeconomic questions, which typically depend on the elasticity of local populations with respect to local changes in utility. We show that both models feature similar short-run population cross-elasticities, in that the elasticity of population in state i with respect to utility in state j is approximately proportional to the gross migration rate between the two states.[5] In other words, if the purpose of a model with

[3]For many questions in the quantitative spatial literature, in which economists are interested in state-level population changes, simulations like these are unnecessary as the aggregate populations can be solved for analytically.

[4]In the SPACE model, the ϵ's represent anything that determines individuals' location choice. While we do not model exactly what these are, we are open to the possibility that this may include information frictions or other frictions that lead to misallocation or which are policy-relevant. We are simply making the point that if the planner could magically relocate a person from one place to another without paying moving costs, that would not be welfare-enhancing in the SPACE model as it would be in the moving cost model.

[5]Even though the models have similar elasticities, the rationale for why the elasticity is related to migration is a bit different. In the SPACE model, the rationale is that migrants are close to indifferent between living in each state, so the mass of people that will move in response to a small shock is close to

migration is to predict short-run effects on populations, both models deliver similar results.

However, in the long run, population cross-elasticities are quite different across the two models. In the SPACE model, population elasticities are the same in the short-run and the long-run, meaning that long-run elasticities are still proportional to the migration between the two places. But in the moving cost model, the long-run population elasticities are approximately the same as a static discrete choice logit model, i.e. the elasticity is proportional to the population share of the shocked region.[6] So the SPACE and moving cost models have completely different long-run population elasticities (in the data, population shares and gross migration rates have very low correlation). This is a problem if we wish to make predictions about long-run populations and we use the wrong model.

A fourth difference follows naturally from the previous one: the dynamics of regions' population changes are quite different in the two models. In the SPACE model, the dynamics are simple. In response to a permanent utility change, the population adjusts fully, contemporaneous to the utility shock. But in the moving cost model, the dynamics are relatively slow and can be unintuitive. In that model, a new set of people every period get a sufficiently large enough shock to move, so a permanent utility shock raises the migration rate and population adjusts slowly. Furthermore, because the long-run elasticities are related to population shares, not migration shares, faraway states from the shock adjust particularly slowly, while nearby states are going to adjust quickly and may overshoot.

Finally, the SPACE model and the moving cost model interpret the data very differently in terms of which states have become higher-utility over time. We show that the SPACE model can use standard exact-hat techniques to map observed population changes onto implied utility changes across time, as in the moving cost model. When we use U.S. population data to infer which states are gaining in terms of relative utility, we get very different answers depending on which model we use. This is critical if we want to estimate the role of policy or economic shocks on welfare.

We finish the paper by discussing the importance of the differences between the SPACE and moving cost model in the context of the literature. We show how the differences we highlight are central to some of the questions that are asked in the dynamic spatial

proportional. In the moving cost model, the rationale is that the extreme value function has a functional form such that the number of people on the margin is proportional to the number of people who make that choice.

[6]So in the moving cost model, the long-run population elasticity of state i and state j to a shock in state k are the same, no matter how different the migration rates between i and k versus j and k are.

literature. We discuss whether various approaches to enrich the moving cost model would deliver similar results to the SPACE model.

In order to clarify the contributions of the model, we wish to be specific regarding the difference between moving costs and persistence in match-specific utility. While mathematically, it is straightforward to specify them in a model (as we will do here), it is important to understand what each term means when we try to map it onto the real world. A typical moving cost model has a one-time irreversible cost borne by people who leave one area for another. In contrast, persistent match-specific utility means that the change in utility when a person moves from one location to the other is both persistent over time and partially reversible should the person move back to the original location. A moving truck and the psychological cost of throwing a goodbye party clearly are moving costs. But many things described as "costs" in the literature are easily reversible—although it may decay with time—and not one-time. Having to live a long way from your friends or favorite amenity is easily reversible and is borne continuously. So even though those are often called "moving costs" in the literature, we think that that terminology is used because existing models have not been able to distinguish persistence in match-specific utility from moving costs. The rest of this paper will give many reasons why this distinction is important.

Literature

How people choose where to live is a classic question in the urban economics literature. Many urban models assume utility is equalized across space in the tradition of Rosen (1979) and Roback (1982). Other more quantitative models assume a discrete choice framework for locations to answer a variety of questions, such as the role of endogenous amenities on location choice (Diamond, 2016) or spatial misallocation on aggregate output (Hsieh and Moretti, 2019).

An increasingly large part of this literature has explicitly looked at the dynamics of location choice, i.e. migration. Since at least Blanchard and Katz (1992), migration has been recognized as a key feature to how regions adjust to economic shocks. In this vein, papers studying the rise or decline of regional economies put a significant emphasis on migration (Caliendo, Dvorkin and Parro, 2019; Allen and Donaldson, 2020; Morris-Levenson and Prato, 2022), and especially the speed at which migration happens (Glaeser and Gyourko, 2005; Kleinman, Liu and Redding, 2023; Amior and Manning, 2018; Davis, Fisher and Veracierto, 2021). Similarly, when aggregating up to the macroeconomy, migration plays a critical role in how quickly countries adapt to changing technologies or external shocks (Tombe and Zhu, 2019; Hao, Sun, Tombe and Zhu, 2020; Eckert and Peters, 2018; Giannone, 2017; Heise and Porzio, 2021; Bryan and Morten, 2019). A growing literature has emphasized how migration, including internal migration, plays an important role in adapting to global warming (Rudik, Lyn, Tan and Ortiz-Bobea, 2021; Cruz and Rossi-Hansberg, 2021; Oliveira and Pereda, 2020). Migration is also known to be an important margin when analyzing housing markets in particular (Schubert, 2021).[7] Central to many of these questions is how elastic is the population of a region to various shocks, over various time horizons. One of the contributions of this paper is to examine how robust those conclusions are to alternative ways of modeling migration.

Corresponding to the growth of interesting questions related to migration, there have also been advances in the ways to model migration. Kennan and Walker (2011) wrote down the canonical model of migration using the dynamic logit formulation. Kaplan and Schulhofer-Wohl (2017), Giannone et al. (2020), Porcher (2020), Mangum and Coate (2019), Zerecero (2021), and Monras (2018) have built on this formulation to include additional realistic features of moving, such as richer information frictions, wealth of migrants, home bias, and nested decision making.[8] Other approaches, such as Coen-Pirani (2010) and Davis et al. (2021) do not use the dynamic logit, but have similar discrete choice models that improve the tractability in a way specific to their goals. All of these models use moving costs to explain the low rates of migration, and potentially adjust those moving costs to explain the high rates of return migration. In contrast, only one paper to our knowledge uses persistence in unobservable match-specific utility to explain low migration rates: Bayer and Juessen (2012). However, the model Bayer and Juessen (2012) is too complicated to extend beyond two regions, limiting its use in many empirical applications.

One type of persistent match-specific utility has been introduced by Zabek (2020),

[7] Howard and Liebersohn (2021) and Howard, Liebersohn and Ozimek (2023) also study the effects of changing location choice on housing markets, but model location choice in a static discrete choice framework rather than explicitly having a notion of migration.

[8] In particular, Kaplan and Schulhofer-Wohl (2017) argues that changes in information frictions can help explain the decline in interstate migration, along with decreases in the different returns to various occupations across space. Giannone et al. (2020) builds a rich model of migration that incorporates wealth and borrowing, to analyze how credit and savings can affect if and where people choose to move. Porcher (2020) builds a tractable model of rational inattention in the dynamic migration context to argue that information frictions are one of the main reasons people do not move. Mangum and Coate (2019) includes both a bias for living near a birthplace, as well as attachment to a place that grows over time spent there, and uses that to argue that shift of the American population to the West and to the South is responsible for slowing labor mobility. Zerecero (2021) also examines a model that includes a preference for birthplace. Monras (2018) looks at the asymmetric response of inmigration and outmigration to local shocks, and builds a dynamic nested logit model to better understand the phenomenon.

Mangum and Coate (2019) and Zerecero (2021), by adding in a preference for living in one's birthplace.[9] These models have some similarity with the SPACE model, in that they also tend to feature smaller moving costs (Zerecero, 2021) and would intuitively feature more return migration than the standard moving cost model. However, adding birthplace preferences to a moving cost model does not generate the square root fact and does not change many of the implications we highlight as being distinct between the two models.

At the same time that models of internal migration have become more popular, empirical evidence focused on the causes and barriers to migration has also grown. For example, Saks and Wozniak (2011) shows migration is cyclical; Kleemans (2015) studies the income shocks that cause migration; Farrokhi and Jinkins (2021) examines the attachment hypothesis using a policy change amongst Danish refugees; Koşar, Ransom and Van der Klaauw (2021) uses a survey experiment to better understand how people make location choice decisions; and Fujiwara, Morales and Porcher (2022) proposes a methodology for uncovering information frictions in location choice. Our paper also contributes to this literature by establishing the additional stylized fact that the t-year migration rate is proportional to $\sqrt{t}$.

1 Data

Throughout the paper, we primarily use credit data from one of the leading credit report providers to measure migration (Gies Consumer and Small Business Credit Panel, 2004-2018), although when we can, we verify the data using the IRS migration data, the American Community Survey, or the Panel Survey of Income Dynamics (IRS Migration Data, 2004-2018; Ruggles, Genadek, Goeken, Grover and Sobek, 2015; Panel Survey of Income Dynamics, 1969-1997).[10] The credit data is a 15-year panel of individuals making up a 1 percent sample of the United States. It records the state of residence in each year, allowing us to calculate migration rates at longer horizons. The GCCP has a similar one-year migration rate to the IRS data, which can be seen in Figure 1. One reason the credit data may have a higher migration rate is that not everyone has a credit report. In particular, lower income people tend to not have credit reports, and are also less likely to

[9] The canonical model in Kennan and Walker (2011) also includes a premium for birthplace.

[10] For other papers using the Gies Consumer and Small Business Credit Panel, see Fonseca (2022), Fonseca and Wang (2022), and Han (2022). DeWaard, Johnson and Whitaker (2019) analyzes a similar credit dataset (the Federal Reserve Bank of New York/Equifax Consumer Credit Panel) on how it can be used to study migration.

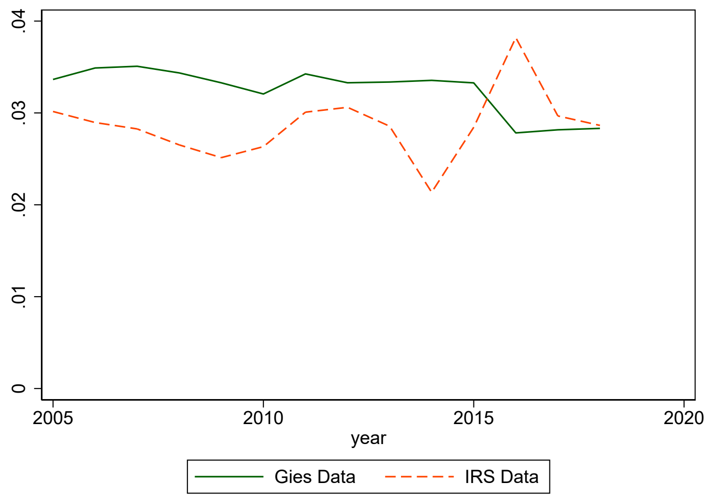


Figure 1: Comparison of 1-year interstate migration rates in IRS and GCCP data.

move.[11]

The data is consistent with two well-known facts that are often the motivation of standard moving cost models. The first fact is that interstate migration is rare. In Figure 1, we show that the GCCP records slightly more than 3 percent of Americans moving between states in any given year. This is slightly higher than the IRS data, which is also shown in the figure. The vast majority of Americans do not move between states in any given year.

The second fact is that migration follows a gravity pattern, meaning that the amount of migration between two states is increasing in each state's population, and decreasing in the distance between them. In Table 1, we show the results from a psuedo-poisson maximum likelihood regression in which we regress migration on log distance and the log populations of the origin and destination states (Silva and Tenreyro, 2006; Correia,

[11]While there are some well-known drawbacks to the IRS data, e.g. it is based only on tax filers, it is one of the most comprehensive administrative datasets keeping track of migration. It is not well understood why the migration rate is so low in 2014 or so high in 2016, as these anomalous values did not show up in other datasets measuring migration (see DeWaard, Hauer, Fussell, Curtis, Whitaker, McConnell, Price and Egan-Robertson (2020)). Similarly, while credit data are not designed as a dataset to study migration, it does have location information, and the bureau gets the addresses from a person's financial accounts. The biggest concern with credit data is that moves may show up with a lag, as people do not always immediately change their addresses with their financial institutions. For our square root fact, we check the robustness to using the Panel Survey of Income Dynamics (1969-1997). The utility of using credit data to discuss internal migration is discussed in depth in DeWaard et al. (2019). The Gies credit data is an unbalanced panel, with yearly observations occurring in May. For matching migration patterns and rates, we focus on the 2004-2005 period, so we only observe data if they had a credit report in both of those years. For some of the dynamics, we address the unbalanced nature of the panel depending on the moments of the data that we are interested in.

Table 1: Gravity Equations

	(1) Migration (IRS)	(2) Migration (Credit)	(3) Migration (IRS)	(4) Migration (Credit)
Log Distance	-0.736***	-0.744***	-1.085***	-1.063***
	(0.0572)	(0.0515)	(0.0694)	(0.0672)
Log Origin Population	0.900***	0.923***		
	(0.0832)	(0.0797)		
Log Destination Population	0.822***	0.893***		
	(0.0976)	(0.0799)		
Observations	2550	2550	2550	2550
Origin and Destination FEs			Yes	Yes

Standard errors are two-way clustered by origin and destination states.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Guimaraes and Zylkin, 2019):

$$\log m_{i \to j} = \alpha \log p_i + \gamma \log p_j + \beta \log \text{distance}_{ij} + \epsilon_{ij} \tag{1}$$

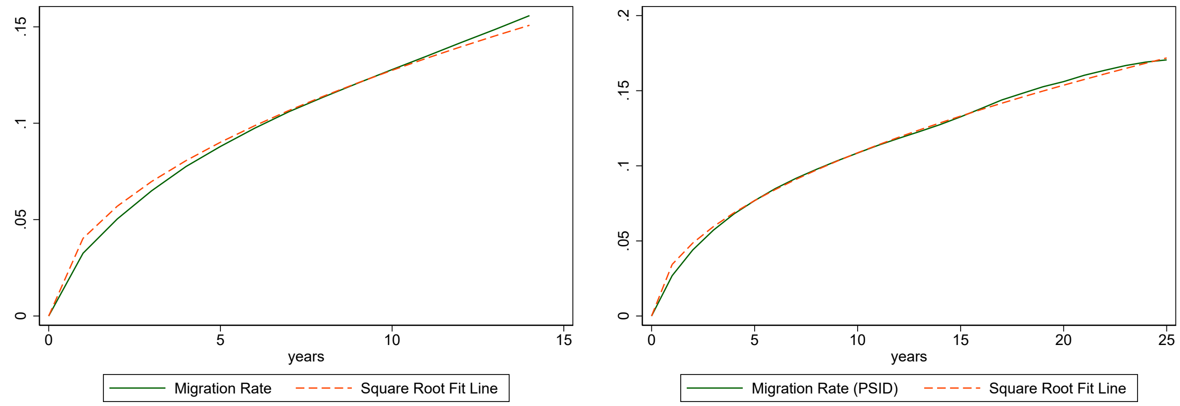
where p_i is the population of i and $m_{i \to j}$ is the 1-year migration from i to j. To calculate distances between states, we use the geographic center of each state from Rogerson (2015) and calculate distances using the formula from Vincenty (1975). We find that both α and γ are positive and β is negative. The coefficients for the IRS data and the GCCP data are similar. We also show a specification with origin and state destination fixed effects, and the coefficients on distance are very similar across the datasets.

2 New Fact

Define the t-year migration rate to be the number of people living in a different state than they did t years ago. The new fact is that the t-year migration rate is approximately proportional to $\sqrt{t}$. In Figure 2(a), the solid line is the t year migration rate in the GCCP. The dashed line is a constant times the square root of t, with the constant chosen to match the level of migration. As is apparent from the figure, the shape of the migration rate is very similar to the square root line.[12]

Of course, since we cannot measure dynamic migration moments in the IRS data, one might wonder if the square root fact is driven by some sort of mismeasurement in the GCCP. In Figure 2(b), we show that the square root fact is also present in data from the

[12]Each point is the mean of a binary variable with millions of observations, so if we tried to put standard errors on the graph, they would not be visible. We show the distribution of the square root fact across state-pairs and across age cohorts in Appendix B.1.



(a) Credit Data (GCCP)

(b) PSID

Figure 2: Migration Rates at Different Horizons. Migration rate at year t is calculated as the percentage of people living in a different state than they did t years ago. Both datasets are unbalanced panels and use any observations in which the state of residence is recorded t years apart. Source: GCCP and PSID.

Panel Survey of Income Dynamics (1969-1997). In fact, we extend the horizon to 25 years and show that it holds through that longer time period as well.[13]

This new fact relates to the more-well-known fact that return migration is common. Many papers in the literature show a significant fraction of workers return to their previous location (e.g. Kennan and Walker, 2011; Kaplan and Schulhofer-Wohl, 2017). One consequence of this fact is that the two-year migration rate is significantly less than twice the one-year migration rate. However, we believe we are the first to document this specific relationship.

2.1 The New Fact is a Puzzle

The square root fact is interesting not only because it is an empirical regularity in need of an explanation, but also because it is at odds with simple existing models. This section shows that the most common model of internal migration in fact leads to a linear relationship between the t-year moving rate and t.

First we outline a standard moving cost migration model. There are a continuum of individuals of mass 1, denoted by n, who can choose to live in locations denoted by i, j,

[13]Since mismeasurement in the GCCP may be a particular concern for young people, we show in Appendix B.1 that it holds for people over 45, where age is estimated by the credit bureau.

or k. An agent that lived in j at $t-1$ has utility:

$$V_{nt}(j) = \max_i \{u_{it} - \delta_{ij} + \epsilon_{int} + \beta \mathbb{E} V_{n,t+1}(i)\} \tag{2}$$

where u_{it} is the common utility for everyone living in i, δ_{ij} is the bilateral moving cost between i and j, and ϵ_{int} is an independent and identically distributed random variable with an extreme-value distribution. We assume ϵ_{int} has a Gumbel distribution with scale parameter 1. If we define $v_{it} \equiv u_{it} + \beta \mathbb{E} V_{n,t+1}(i)$, then the migration probability is given by:

$$\frac{m_{i\to j,t}}{p_{it}} = \frac{e^{-v_{jt}-\delta ij}}{\sum_k e^{-v_{kt}-\delta_{ik}}} \tag{3}$$

What does this model predict for the dynamics of migration, especially for the shape of the t-year migration rate? When moving costs are high—which is required to match the low amounts of interstate migration in the data—then the following proposition shows that the t-year migration rate is approximately linear in t.

To set up the proposition, it is helpful to suppose that moving costs are given by $\delta_{ij} = \delta'_{ij} + \Delta$ when $i \neq j$, and $\delta_{ii} = 0$. For $i \neq j$, migration costs consist of a pair-specific component that governs the relative amount of migration to j, and Δ, a common component which governs the overall amount of migration in the economy. This way, when we change Δ, we are not changing the relative amount of migration from i to j versus i to k.

Proposition 1. *In the steady-state of a moving cost model, as the common component of moving costs go to infinity, the t-year migration is proportional to t.*

$$\lim_{\Delta\to\infty} \frac{m^t_{i\to j}}{m^1_{i\to j}} = t$$

where $m^t_{i\to j}$ is the t-year migration from i to j.

Proofs are collected in Appendix A.

This proposition establishes that the square root fact is not a natural consequence of our standard models. In the standard model, we infer high moving costs based on the fact that migration is low, and this proposition establishes that high moving costs imply a linear relationship between the t-year migration rate and t.

In many calibrated models, moving costs are large but less than infinity, so we also check that the model estimated to the data would generate a relationship that looks linear.[14] First, using the standard moving cost model, we estimate the moving costs that would correspond to the GCCP migration rates in 2004-2005. Then we calculate the within-model t-year migration rate, by simulating the model for millions of observations over 15 years. Figure 3a shows the results of this simulation. The relationship between the t-year migration rate and t looks very linear, and does not do a good job of approximating the data.

In simulations, it is also to calibrate migration costs that depend on variables that the reader might think could break the linear relationship. For example, many models include as a state variable whether the person moved in the previous year and if so, from where (e.g. Kennan and Walker, 2011). Here, we use the GCCP data to calibrate a model in which the moving costs depend on the state of residence, but also the state of residence from the year before. We show the results in Figure 3b. While this additional state variable adds a kink at $t = 1$, the relationship becomes fairly linear again for larger t's.[15]

Similarly, one might think that adding age as a state variable could help match the square root fact since migration rates decline over the life-cycle (through the lens of the standard model, moving costs increase in age). Calibrating the model to depend on the GCCP's age variable still leaves a very linear relationship, as can be seen in Figure 3c.[16]

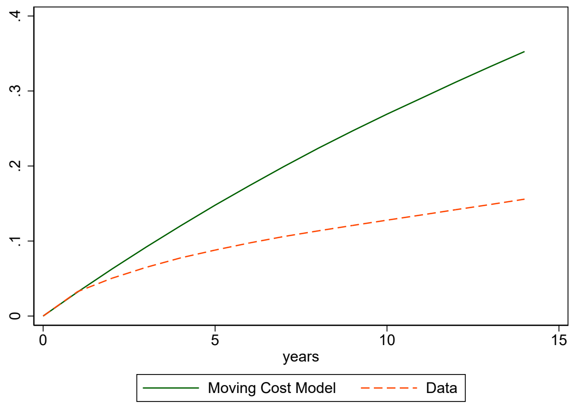
Finally, we also check whether adding the state of birth is helpful to match the curvature of the t-year migration rate, since moving back home may be a high-enough probability event to make Proposition 1 a bad approximation.[17] For this calculation, we use the American Community Survey (ACS) data, which records the state of birth, as well as interstate migration, to calibrate the t-year migration rate. The migration rate in the ACS is modestly lower, so the migration costs will be calibrated to be higher. However,

[14]Alternatively, a reader might wonder if the fact that the world is not in a steady-state could generate a square root pattern. However, as documented in Jia, Molloy, Smith and Wozniak (2023) and other papers, gross migration flows are much larger than net migration flows, often by an order of magnitude, so that cannot explain the difference between moving cost models and the square root fact. In unshown simulations, we estimate the a moving cost model where the moving costs vary year by year to match the migration rates in each year, and they are not appreciably different.

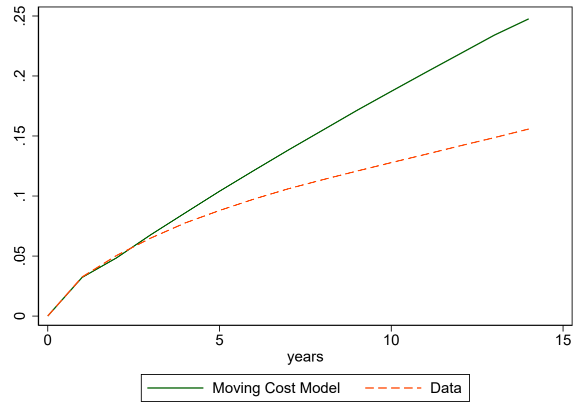
[15]Building on this exercise, one could of course match the square root fact up to 14 years by conditioning the moving costs on the person's location for the last 14 years. But since the model is likely to start making counterfactual predictions about the dynamics after 14 years, one would still need to proceed with caution beyond that horizon.

[16]The GCCP imputes age, so that may introduce errors into the calibration. However, Figure A2 confirms that the GCCP does pick up the lifecycle decline in migration.

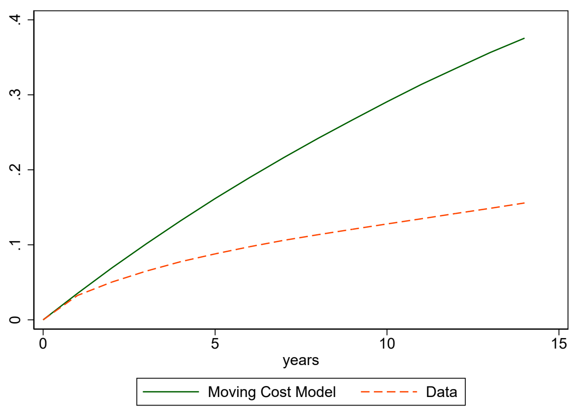
[17]The reader might wonder if other demographics would matter. For example, college educated workers are known to move more (Molloy, Smith and Wozniak, 2011). However, for immutable characteristics, the aggregate migration rate would just be the average of the different groups, so if the groups have linear t-year migration rates, then the aggregate would also be linear. So as long as migration costs are high for every group, then the t-year migration rate will still be linear.



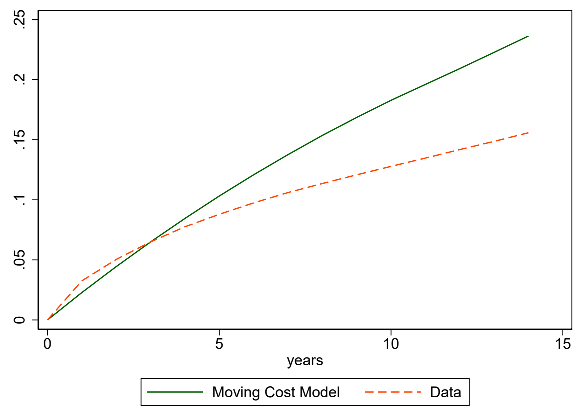
(a) Baseline Moving Cost Model



(b) Previous Year's Location State Variable



(c) Age State Variable



(d) Birthplace State Variable

Figure 3: Do existing moving cost models match the square root fact? Each panel compares the t-year migration rate in simulations of the moving cost model described in Section 2.1 to the data. The model is calibrated by picking migration costs to exactly match interstate migration flows. In Panel (a), migration costs vary by origin-destination pair. In Panel (b), migration costs are allowed to vary by the interaction of location in the year prior, origin, and destination (i.e. someone that had lived in state X then moved to state Y will have different moving costs than someone that lived in state Y for two years). In Panel (c), migration costs vary by the interaction of age, origin, and destination. And in Panel (d), migration costs vary by the interaction of birthplace, origin, and destination. The data is from the GCCP, and for panels (a)-(c), the model is fit on the GCCP data. Panel (d) is fit using ACS data, which has a lower 1-year migration rate, and so the data and model do not match even at $t = 1$.

what we are primarily interested in is the linear relationship, which does not change with this additional state variable.

Of course, a flexible enough model of moving costs could hit the square root fact. For example, one could match many of the dynamic facts about migration by assuming moving costs increase with the time spent in a particular location and that moving costs are lower when returning to a past location. Similarly, if we let people be of different unosbervable types with lots of heterogeneity in moving costs, we could also generate the square root fact.

We have four comments about this line of thought.

First, these more flexible models are not the models that people actually use. Recent papers that have estimated dynamic effects in spatial models (Caliendo et al., 2019; Kleinman et al., 2023) do not include features that would generate the square root fact. Therefore, much of what the literature knows about the macro dynamics of internal migration is built on models that do not match an important aspect of the micro dynamics.[18]

Second, and related to the first point, these more-flexible models lose a lot of the tractability and economic interpretation of the standard moving cost model. If moving costs depend on a long history of locations, then calculating the macro elasticities depends on keeping track of the size of the population with each of those histories, which is computationally expensive. It is also harder to conceptualize what a moving cost is in the real world when it changes based on tenure, location histories, and unobservable characteristics of a person.

Third, these models can replicate any relationship between the t-year migration rate and t. To hit the square root fact, you have to paramaterize these models just right to not have too much or too little concavity. The model that we propose below can only generate a square root, and it does so without the complexity or the flexibility of a model with lots of moving cost heterogeneity. To use the language of Fudenberg, Gao and Liang (2023), these models are not "restrictive" compared to the model presented below.[19]

Finally, even if the previous three points are not sufficient to convince the reader that they might prefer the model in this paper to a very-flexible moving cost model, the reader

[18] People do write down models which have excess return migration in the first period (Kennan and Walker, 2011; Kaplan and Schulhofer-Wohl, 2017). But these are also not flexible to match the square root fact.

[19] Fudenberg et al. (2023) explains why more restrictive models are desirable compare to less restrictive: "A potential reason for this preference is that models are often meant to capture behavior in related by not-identical domains. Given enough data, models that are very unrestrictive will fit any specific data set well, but may do so by learning idiosyncratic details of those datasets that do not in fact transfer across settings. In contrast, if a highly specific and structured model happens to fit a data set well, this may generate more confidence that the model's structure extends to other settings."

might still be interested in knowing about a new model can fit the micro data equally well and makes different predictions for welfare and counterfactuals. This should inform us about the robustness of the conclusions drawn with the standard model.

3 New model that can match the new fact

This section introduces a new model of internal migration which can resolve the square root puzzle from the previous section. Rather than depend on moving costs, it assumes that the match-specific utility (the ϵ's) are spatially-correlated and persistent. In fact, because the model features Spatially and Persistently Auto-Correlated Epsilons, we call it the SPACE model.

As in the moving cost model, there is a continuum of individuals with mass 1, a finite number of discrete locations, and discrete time. We keep the same notation where n denotes the individual, i the location, and t the year. Individuals pick their location to maximize utility:

$$V_{nt}(\vec{\epsilon}_{nt}) = \max_i \{u_{it} + \epsilon_{nit}\} + \beta \mathbb{E}[V_{n,t+1}(\vec{\epsilon}_{t+1})|\vec{\epsilon}_t] \tag{4}$$

where u_{it} is a common flow utility for location i and ϵ_{nit} (the ith element of vector $\vec{\epsilon}_{nt}$) is a person-location-match-specific utility.[20] Note that the choice of location i does not affect the continuation value because there are no moving costs, so the choice is made sequentially each period, and it has no effect on future choices.

To generate spatial correlation, we assume that $\epsilon_{nt} \equiv (\epsilon_{1nt}, ..., \epsilon_{Int})$ is distributed as a generalized extreme value distribution, where the marginal distribution of ϵ_{int} is a Gumbel distribution, but they are not necessarily independent of one another:[21]

$$\epsilon_{nt} \sim F(\cdot)$$

where

$$F(\epsilon_{1nt}, ..., \epsilon_{Int}) = \exp(-G(e^{-\epsilon_{1nt}}, ..., e^{-\epsilon_{Int}})) \tag{5}$$

where G is a correlation function in the sense of McFadden (1977). To be specific, G is

[20] We do not take a stand on where the u_i's originate, so the reader can think of the SPACE model as being the migration block of a spatial model, and that the u_i's would originate in the housing, production, and amenities blocks.

Note that because there are no moving costs, the continuation value $v_{it} = u_{it} + \beta V_{n,t+1}(\vec{\epsilon}_{t+1})$ differs from u_i by just a constant.

[21] See Lind and Ramondo (2023) as an example of a generalized extreme value distribution in trade.

defined over the range of N non-negative real numbers, and it must satisfy the following properties: it is non-negative; it is homogenous of degree 1; the limit when any one of its arguments approaches infinity is infinity; and the cross-partial with respect to any k distinct arguments is nonnegative if k is odd and nonpositive if k is even.

Under these assumptions, the choice probability of an agent choosing choice i is:

$$p_i = e^{v_i} \frac{G_i(e^{v_1}, ..., e^{v_I})}{G(e^{v_1}, ..., e^{v_I})}$$

where G_i is the partial derivative of G with respect to its ith argument. See McFadden (1977) for the derivation.

We also wish to make the ϵ correlated not just over space but also over time. To do that, we assume that the joint distribution of ϵ_{nt} and ϵ_{nt+1} is given by:

$$(\epsilon_{nt}, \epsilon_{nt+1}) \sim F_2(\cdot, \cdot)$$

where

$$F_2(\epsilon_{1nt}, ..., \epsilon_{Int}, \epsilon_{1nt+1}, ..., \epsilon_{Int+1}) = \exp\Big(- G\big(H(e^{-\epsilon_{1nt}}, e^{-\epsilon_{1nt+1}}), ..., H(e^{-\epsilon_{Int}}, e^{-\epsilon_{Int+1}})\big)\Big)$$

and

$$H(x_1, x_2) = \left(x_1^{\frac{1}{1-\rho}} + x_2^{\frac{1}{1-\rho}} \right)^{1-\rho}$$

where $\rho < 1$. We will further assume that $G(H(\cdot,\cdot), ..., H(\cdot,\cdot))$ is also a correlation function under the criteria above.

If G has a cross-nested structure with exponents γ_k for each nest k, then as long as $\rho > \gamma_k$ for all k, G will be a correlation function. This claim is true because under those conditions, G is the sum of several nested logit correlation functions, and the sum of correlation functions is a correlation function. This claim is important because then the function F_2 is indeed a proper cumulative distribution function. Lind and Ramondo (2023) prove that cross-nested formulations of G can approximate any correlation function, so combined with this proposition, then any correlation function can be approximated by a correlation function that permits some ρ that induces the persistence we desire to add to the model.

Note that the cumulative distribution function of ϵ_{nt} can be calculated by taking the limit as ϵ_{nt+1} goes to infinity. In this case, $\lim_{\epsilon_{t+1}\to\infty} H(e^{-\epsilon_t}, e^{-\epsilon_{t+1}}) = e^{-\epsilon_t}$. So the marginal distribution of ϵ_{nt} is given by F from equation (5). A similar argument applies so that the marginal distribution of ϵ_{nt+1} is also given by F. Hence, the cross-sectional distribution of ϵ_i is time-invariant, even though an individual's ϵ_{nt} will not be the same as ϵ_{nt+1}.

The joint distribution F_2 implies a conditional distribution $\tilde{F}(\epsilon_{nt+1}|\epsilon_{nt})$. This can be iterated as a Markov chain to calculate distributions of future ϵ's.

Migration occurs when the locations i that maximize $u_{it} + \epsilon_{nit}$ and $u_{i,t+1} + \epsilon_{ni,t+1}$ are different.

Proposition 2. *Under these assumptions and when the v_i's are fixed, then migration from i to j, is given by:*

$$m_{i \to j} = (1-\rho)e^{v_i+v_j}\left(\frac{G_i(e^{v_1},...,e^{v_I})G_j(e^{v_1},...,e^{v_I})}{G(e^{v_1},...,e^{v_I})^2} - \frac{G_{ij}(e^{v_1},...,e^{v_I})}{G(e^{v_1},...,e^{v_I})}\right) \tag{6}$$

Alternatively,

$$m_{i \to j} = (1-\rho)p_i p_j \left(1 + \tau_{ij}(e^{v_1},...,e^{v_I})\right)$$

where $\tau_{ij} = -\frac{G_{ij}(e^{v_1},...,e^{v_I})G(e^{v_1},...,e^{v_I})}{G_i(e^{v_1},...,e^{v_I})G_j(e^{v_1},...,e^{v_I})}$.

Please refer to Appendix A for the proof.

This formulation resembles a gravity equation. τ_{ij}, which is non-negative, denotes the correlation between two alternatives, and more correlated alternatives will have more migration between them. Similarly, as migration is more correlated over time (higher ρ), there will be less migration.

Corollary 1. *The cross-elasticity of population to utility is given by the migration rate times a constant.*

$$\frac{\partial p_i}{\partial v_j} = \frac{1}{1-\rho} m_{i \to j}$$

when $i \neq j$.

See Appendix A for the proof. This corollary is important because in many applications, we are interested in the population elasticity to local shocks. This corollary tells us that migration is a sufficient statistic to know these elasticities up to a constant.

The migration in Corollary 1 is a steady-state migration between i and j, which in the steady-state of the model is equal to the migration from j to i. In practice, migration is rarely exactly balanced. Our recommendation is to use the logarithmic average, i.e. $(m_{i \to j} - m_{j \to i})/\log(m_{i \to j}/m_{j \to i})$, because in a parametric version of the model that we introduce in the next section, this is a very precise approximation of the steady-state migration we need for corollary 1. See Appendix C for details.

A reader may wonder if the specific features of the SPACE model are "cooked up" to match facts but lack a basis in reality. Rather, we think that the SPACE model has two very realistic features of preferences. First is that the match-specific utility for location is persistent over time. People mostly cite family and work in surveys about why they move (Jia et al., 2023). People's feelings about these things are surely correlated over time, and it is an empirical fact that each of these things is persistent in terms of location. Second is that match-specific utility is spatially-correlated. Again, when you think about people's stated preferences, the ability to live near family is highly-correlated across space. If state i is close to family, then states near i are also close to family. Industrial composition, i.e. the types of jobs people can get, also tend to be geographically concentrated. Natural amenities or regional cultures—other possible sources of match-specific utility—are also spatially correlated. The functional forms of equations are indeed convenient mathematically and are likely not precisely true, but it should be hard to argue that spatially and auto-correlated match-specific utility is somehow less realistic than the i.i.d. utilities of a moving cost model.

Having setup the model, we now present a proposition analogous to Proposition 1 from the moving cost model, to see whether the SPACE model can match the square root fact. In this case, the limit we consider that makes it so that there is only a little migration is for the persistence parameter ρ to approach 1.

When ρ is close to 1, ϵ_{int} will resemble a random-walk with logistic innovations. It is well known that the standard deviation of a random walk grows with the square root of time. As the standard deviation grows, the odds of crossing a threshold—which corresponds to moving in this model—grow roughly proportionally to the standard deviation, hence generating the square root fact. We formalize this intuition in the following lemma:

Lemma 1. *Define Λ^n to be the convolution of n i.i.d. mean-zero logistic random variables. Then,*

$$\lim_{\rho \to 1} \frac{m^t_{i\to j}}{m^1_{i\to j}} = \frac{\mathbb{E}[|\Lambda^{2t}|]}{\mathbb{E}[|\Lambda^2|]}$$

where $\mathbb{E}[|\cdot|]$ denotes the mean absolute deviation.

Please refer to Appendix A for the proof. This proposition establishes that for ρ close to 1, relative migration over t years will be proportional to the mean absolute deviation of the convolution of $2t$ independent logistic random variables, following the intuition from above. Importantly, we can calculate bounds on this ratio in the following corollary:

Proposition 3. *As $\rho \to 1$, the ratio of t-year migration to 1-year migration for any state pair, is bounded below by $\sqrt{t}$ and bounded above by $\sqrt{\pi/3}\sqrt{t}$:*

$$\sqrt{t} \leq \lim_{\rho \to 1} \frac{m^t_{i \to j}}{m^1_{i \to j}} \leq \sqrt{\frac{\pi}{3}}\sqrt{t}$$

Please refer to Appendix A for the proof. $\sqrt{\pi/3}$ is approximately 1.023, so the bound is tight. This establishes that the migration rate is within 2.3 percent of the square root fact that motivated the model, when ρ is close to 1.[22]

3.1 Calibration

Up to this point, we have not taken a stand on the functional form of G because there are many possible G that would exactly match the populations and migration in the data. Of course, in many settings where the outcome of interest is the change in populations, the specific choice of G makes no first-order difference, due to the previous result that $\frac{\partial p_i}{\partial v_j} = \frac{1}{1-\rho} m_{i \to j}$.

Nonetheless, in some applications, it will be helpful to take a stand on the specific functional form of G. For most of these applications, it is helpful to pick a functional form that allows for simple formulas for population changes, and the ability to simulate an individual's migration path in a computationally easy way.[23] We will also be interested in finding a G that is consistent with a value of ρ near 1, because that is the parameterization of ρ that leads to the square root fact.

Consider the following G functional form, which generates a cross-nested logit with up to two locations per nest:

$$G = \sum_i w_i x_i + \sum_i \sum_{j<i} \left((w_{ij} x_i)^{\frac{1}{1-\gamma}} + (w_{ji} x_j)^{\frac{1}{1-\gamma}} \right)^{1-\gamma}$$

Assume we pick the w's such that $G = 1$ at our initial levels of $x_i = e^{v_i}$. Then, the initial

[22] A tight band in Proposition 3 is intuitive because, if the innovations were normally distributed and Lemma 1 referenced normal distributions instead of logistic ones, the ratio would be exactly the square root, without the 2.3 percent deviation. Similarly, if the lemma were based on the standard deviation rather than the mean absolute deviation, the ratio would also be precisely the square root. Although this exact relationship does not hold for the MAD of logistic distributions, logistics are approximately normal, and the MAD is approximately proportional to the standard deviation for many distribution families. Consequently, it is not surprising that the relationship holds approximately.

[23] A calibration also allows us to solve for migration outside of steady-state, which we do in Appendix C.

population will be given by

$$p_i = w_i e^{v_i} + \sum_{j \neq i} \left((w_{ij} e^{v_i})^{\frac{1}{1-\gamma}} + (w_{ji} e^{v_j})^{\frac{1}{1-\gamma}} \right)^{-\gamma} (w_{ij} e^{v_i})^{\frac{1}{1-\gamma}}$$

and the initial migration by

$$m_{ij} = (1-\rho) \left(p_i p_j + \frac{\gamma}{1-\gamma} \left((w_{ij} e^{v_i})^{\frac{1}{1-\gamma}} + (w_{ji} e^{v_j})^{\frac{1}{1-\gamma}} \right)^{-\gamma-1} (w_{ij} e^{v_i})^{\frac{1}{1-\gamma}} (w_{ji} e^{v_j})^{\frac{1}{1-\gamma}} \right)$$

This choice of functional form leads to a cross-nested logit, in which outer nests contains either one or two locations. The within-nest elasticity of substitution is $\frac{1}{1-\gamma}$ and the across-nest elasticity of substitution is 1.

In order to do micro-simulations, it will be helpful to parameterize γ and w_{ij} as a function of ρ, and consider the limit as $\rho \to 1$. In particular, we will parameterize $1 - \gamma = \frac{1-\rho}{1-\tilde{\rho}}$ where $\tilde{\rho}$ is a constant, so as $\rho \to 1$, $\gamma \to 1$ as well. This will allow us to consider the nest for each person as fixed over time, but to allow for agents to move within nests. In order to approximately match the square root fact, we will try to choose a $\tilde{\rho}$ as close to 1 as possible while also matching migration rates in the data.

Rather than calibrate the w_{ij}, we will calibrate

$$\tilde{w}_{ij} \equiv (w_{ij} e^{v_i})^{\frac{1}{1-\gamma}} \left((w_{ij} e^{v_i})^{\frac{1}{1-\gamma}} + (w_{ji} e^{v_j})^{\frac{1}{1-\gamma}} \right)^{-\gamma}$$

Note that this is also a function of ρ since it depends on γ. From there, the w_{ij}'s can be backed out as $w_{ij} = e^{-v_i} \tilde{w}_{ij}^{1-\gamma} (\tilde{w}_{ij} + \tilde{w}_{ji})^{\gamma}$. We will also calibrate $w_i = 0$ because it allows us to have a higher $\tilde{\rho}$ if there are no agents that will never move. Under this assumption,

$$p_i = \sum_{j \neq i} \tilde{w}_{ij} \tag{7}$$

and $\frac{1}{1-\rho} m_{ij} = p_i p_j + \frac{\gamma}{1-\gamma} \frac{\tilde{w}_{ij} \tilde{w}_{ji}}{\tilde{w}_{ij} + \tilde{w}_{ji}}$. But as $\rho \to 1$ and holding constant $\frac{1-\rho}{1-\gamma} \equiv 1 - \tilde{\rho}$, then we can rewrite this formula as

$$\frac{1-\tilde{\rho}}{m_{ij}} = \frac{1}{\tilde{w}_{ij}} + \frac{1}{\tilde{w}_{ji}} \tag{8}$$

We want to maximize $\tilde{\rho}$ subject to (7) and (8). Call the Lagrange multipliers on (7), λ_i and the ones on (8), ν_{ij}. The first-order conditions with respect to $\tilde{w}_{ij}$ and $\tilde{w}_{ji}$, which are necessary for it to be a maximum are $\lambda_i = \frac{1}{\tilde{w}_{ij}^2} \nu_{ij}$ and $\lambda_j = \frac{1}{\tilde{w}_{ji}^2} \nu_{ij}$. So $\frac{\tilde{w}_{ij}}{\tilde{w}_{ji}} = \sqrt{\frac{\lambda_j}{\lambda_i}}$. We can use (8), along with this new formulation to solve for the $\tilde{w}_{ij}$ in terms of these

Lagrange multipliers:

$$\tilde{w}_{ij} = \frac{1}{1-\tilde{\rho}} m_{ij} \left(1 + \sqrt{\frac{\lambda_j}{\lambda_i}}\right) \tag{9}$$

Plugging (9) into (7) and multiplying by $\sqrt{\lambda_i}$,

$$(1-\tilde{\rho})\sqrt{\lambda_i} = \sum_{j \neq i} \frac{m_{ij}}{p_i} \left(\sqrt{\lambda_i} + \sqrt{\lambda_j}\right)$$

If we define the $I \times I$ matrix M as

$$M_{ij} = \begin{cases} \frac{m_{ij}}{p_i} & \text{if } i \neq j \\ \sum_{k \neq i} \frac{m_{ik}}{p_i} & \text{if } i = j \end{cases}$$

Then

$$(1-\tilde{\rho})\ell = M\ell$$

where ℓ is an $I \times 1$ vector and $\ell_i = \sqrt{\lambda_i}$.[24]

This equation is satisfied for any eigenvector ℓ of matrix M, and corresponding eigenvalue $(1-\tilde{\rho})$. However, we also require that both ℓ and $(1-\tilde{\rho})$ are positive. When $m_{ij} > 0$, the matrix has all positive values, so there exists a unique real eigenvalue corresponding to an eigenvector with all positive entries by the Perron (1907)-Frobenius (1912) theorem. That eigenvalue is our calibration for $(1-\tilde{\rho})$, and the corresponding eigenvector tells us $\tilde{w}_{ij}$ and therefore w_{ij}. In particular with the GCCP migration data, $\tilde{\rho}$ is equal to .8913, which can be used to simulate migration.

This calibration is useful for two reasons. First, because it leads to simple formula for populations and straightforward exact-hat algebra. Define $\hat{p}_i = p_i'/p_i$ and $\hat{v}_i = \exp(\frac{v_i'-v_i}{1-\rho})$. Then

$$\hat{p}_i = \hat{v}_i^{1-\tilde{\rho}} \sum_{j \neq i} \frac{\tilde{w}_{ij}}{p_i} \frac{(\tilde{w}_{ij} + \tilde{w}_{ji})}{\tilde{w}_{ij}\hat{v}_i^{1-\tilde{\rho}} + \tilde{w}_{ji}\hat{v}_j^{1-\tilde{\rho}}}$$

The derivation can be found in Appendix A. Recall that $\tilde{w}_{ij}$ are easily calculable in equation (9) from the eigenvector ℓ and data on migration.[25]

Second, because each person is drawn into choosing between two locations, and never

[24] The matrix M is based on data which is typically not in a steady-state. For example, we use data on migration based on locations in 2004 and 2005, and generally, $m_{i \to j} \neq m_{j \to i}$. For m_{ij}, we recommend using the logarithmic average of $m_{i \to j}$ and $m_{j \to i}$ based on our analysis in Appendix C. For p_i, we use the populations from 2005.

[25] If a reader wishes to calibrate a SPACE model to do exact hat algebra without having to find an eigenvector, a more straightforward alternative is to calibrate $\tilde{w}_{ij} = \tilde{w}_{ji} = \frac{2m_{ij}}{1-\tilde{\rho}}$, and $\tilde{w}_i \equiv w_i e^{v_i} =$

considers anywhere else. Their actions are as if they were drawing two Gumbel-distributed ϵ's that are independent between the two locations, but correlated over time with correlation parameter $\tilde{\rho}$.

In particular, it is computationally cheap to simulate sequences of conditional Gumbels where ϵ_t and ϵ_{t+1} have joint distribution

$$\exp(-(e^{-\epsilon_{it}/(1-\tilde{\rho})} + e^{-\epsilon_{it+1}/(1-\tilde{\rho})})^{1-\tilde{\rho}})$$

because there is a closed-form formulation of the CDF of ϵ_{t+1} given ϵ_t. We can easily draw a sequence for location i and another sequence for location j, and then that person will pick i when $\epsilon_{it} > \epsilon_{jt} + \log \frac{\tilde{w}_{ij}}{\tilde{w}_{ji}}$

The purpose of a calibration that easily allows for simulation is to verify that the model matches the dynamic moments of the data besides the square root fact. It is also helpful to verify that even for $\tilde{\rho} < 1$, the approximation of the square root fact is still reasonable.

3.2 Dynamics of Migration in the Calibrated SPACE model

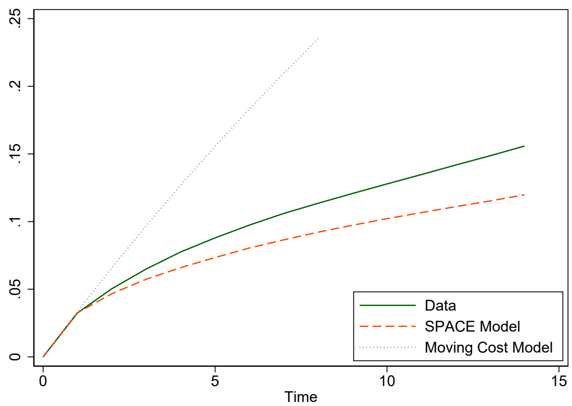
In Figure 4a, we show the t-year migration rate does follow a square root pattern in both the data and in a simulation of the SPACE model. For the data, we include any observations for which we have credit reports t years apart, so it should be noted that the sample changes slightly depending on t.[26] The model does not match the data perfectly, with the lines diverging over time. In part, this is because the model is calibrated to match the one-year migration rate, and the fourteen-year migration rate, which in the simulation is about $\sqrt{14}$ times the one-year migration rate, is going to be sensitive to that choice. The figure also presents the same exercise for a simulation of the moving cost model, which is much more linear and diverges much more from the data.

$p_i - \sum_{j \neq i} \tilde{w}_{ij}$. Under those assumptions, the exact hat algebra is given by:

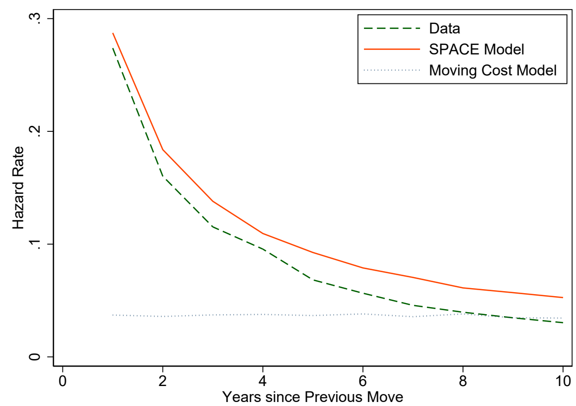
$$\hat{p}_i = 1 - \frac{2}{1-\tilde{\rho}} \frac{m_i}{p_i} + \sum_{j \neq i} \frac{4}{1-\tilde{\rho}} \frac{m_{ij}}{p_i} \frac{v_i^{1-\tilde{\rho}}}{\hat{v}_i^{1-\tilde{\rho}} + v_j^{1-\tilde{\rho}}}$$

This is a simpler calibration because it does not require taking eigenvalues and eigenvectors, but the upper bound on $\tilde{\rho}$ is $\max_i\{1 - 2\frac{m_i}{p_i}\}$, which is in practice a bit less than 0.8, so the approximation of ρ close to 1, which generates the square root fact, is less precise. In simulations, this generates extra concavity compared to the square root fact.

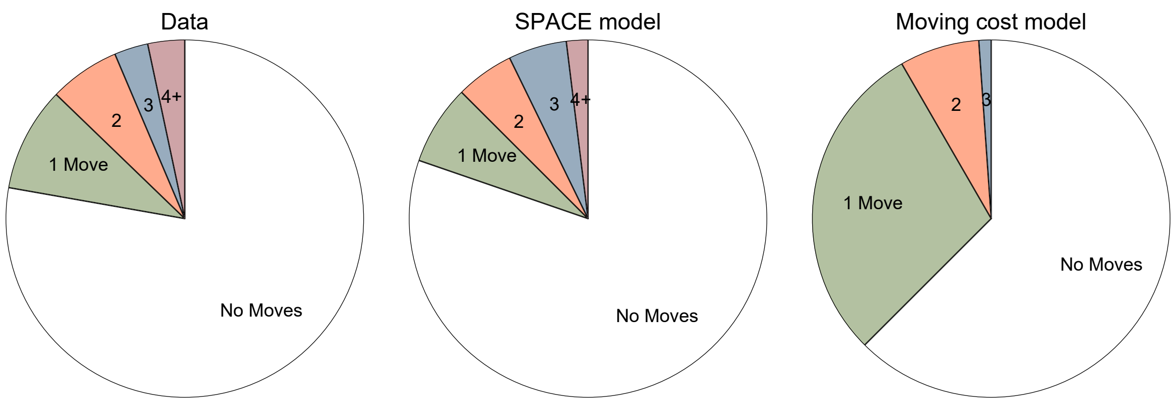
[26] Focusing only on a balanced panel of individuals gives an indistinguishable pattern, but raises concerns about excluding younger people who are most likely to move. Each of the dynamic moments that we look at in Figure 4 selects a slightly different sample of people, so the fact that the model is still



(a) Migration rate in the data and models



(b) Hazard Rate of migration



(c) Number of moves in 14 years

Figure 4: Dynamic moments. In panel (a), the t-year migration rate is calculated as the percent of people living in a different state than they were t years ago. Data is from an unbalanced panel, and included any observations from 2004-2018 for which the state of residence is observed t years apart. In panel (b), the conditional probability of migration is plotted. For the value at x years, the probability of migration is conditional on the person having migrated x years previously and remained in the same state ever since. It is broken up into return migration, which is when the person moves back to the original state, and onward migration if they move to a third state. In panel (c), the number of moves in 14 years is calculated for people whose state is observed in every year from 2004-2018.

Of course, the t-year migration rate is not the typical way the dynamic moments of migration are presented in the data, so it is interesting whether the model is able to capture the more-commonly-examined moments as well. A natural moment is the conditional probability of moving given a previous move, i.e. the hazard rate of migration.

Figure 4b shows the probability of another migration at different time horizons after an interstate move.[27] Since we do not target these statistics in the calibration, the simulated statistics do not match the data perfectly. But the general pattern is similar, especially its decay as the person has lived in the state for longer.

The intuition for the decreasing hazard rate over time is simple. Conditional on having moved recently, the agents are likely relatively indifferent between the two regions, and are likely to move back. The longer they have stayed in one region, the more likely that their accumulated utility shocks have drawn them further away from being indifferent, so the probability of migration decreases over time. The literature has typically focused on the concept of "attachment" to explain this phenomenon (Mangum and Coate, 2019; Farrokhi and Jinkins, 2021). In the SPACE model, people who have lived in a location for longer are more attached, but it is because their repeated decision not to move has revealed that they like the location, not due to an economic force that increases their utility by staying there longer.

Another easy-to-measure moment is the distribution of the number of interstate moves over time. Figure 4c looks at how many moves are made over a 14 year period. In the data, a large majority of people make zero moves, but some people make many moves. Here, we include in this chart only people for whom we have data in all 15 years (for up to 14 possible moves). The model is able to capture the large fraction of people that never move, as well as come close to the data on the number of people that move once or twice. Importantly, it captures the fact that a few percent of people move four or more times over the fourteen years. The figure also includes similar statistics for a moving cost model, which does a much worse job.

4 Does the new model matter?

The previous section introduced a new model and showed that it did a better job at matching dynamic moments in the micro data. In this section, we explore the implications

a good match across the different moments shows the sample does not matter that much.

[27]To be included in this analysis, a person must show up for the number of years that would be necessary to calculate the statistic, but we do not use a balanced panel.

of that model, especially by comparing it to the workhorse model. For some questions, we find that the differences are minimal, while for others we think there are substantial differences.

4.1 Micro Forecasting

The first reason we might care about the difference between the two models is for purposes of forecasting the location of an individual agent. Suppose we observe the agent's location in 2004, and wish to forecast where they will live in every subsequent year until 2018. We use the calibrated versions of each model to do the forecasting. We judge the performance of the models using the mean Kullback-Leibler divergence. Specifically, we simulate each model for ten million people, and then for each initial state, we calculate the simulated probability that a person who was in state i in 2004 ends up in state j in year t. Then using people's true locations in the data, we calculate the relative mean log likelihood,

$$KL_t = \frac{1}{N} \sum_n \log \left(\frac{\mathbb{P}_{\text{data}}(\text{lives in } j \text{ in } t | \text{lived in } i \text{ in } 2004)}{\mathbb{P}_{\text{model}}(\text{lives in } j \text{ in } t | \text{lived in } i \text{ in } 2004)} \right)$$

for each year. We plot KL_t in Figure 5.

Both models initially have a low Kullback-Leibler divergence as they are about equally good at predicting locations in 2005 since they were both parameterized to match the migration data in that year. But over time, the moving cost model's Kullback-Leibler divergence grows sharply, suggesting that the log likelihood is on average about 0.13 log-points worse per observation than the maximum possible performance of any model by 2018. In contrast, the SPACE model has a Kullback-Leibler divergence of 0.015 log-points in 2018, suggesting limited room for further improvement.

The SPACE model does better over time because it can match the dynamic moments. In particular, many more people end up moving away from their initial state in the moving cost model because moving probabilities are independent over time, whereas the SPACE model is better able to match the total number of people who leave.

4.2 Moving costs need not be large

Kennan and Walker (2011) estimate an average moving cost of $312,146 (in 2010 dollars).[28] This is more than six times the median household income in that year, which was $49,445

[28] They also include an analysis of moving costs conditional on moving, but they include the payoff shocks in the moving costs, so find that the average moving cost is actually very negative.

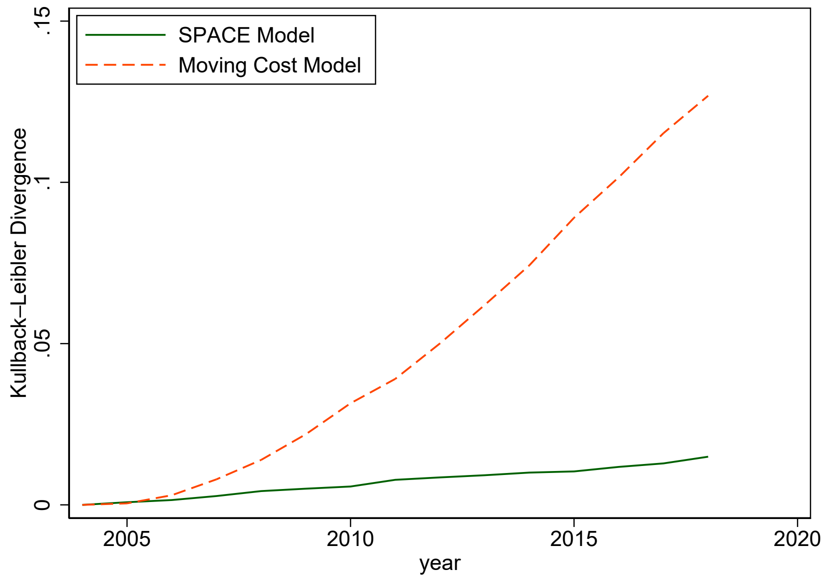


Figure 5: Kullback-Leibler divergence of the SPACE and Moving Cost models

(Census, 2011).[29] Such a large cost is one of the main reasons that most people do not move, in their model. Economists can argue about whether that number is reasonable, and even within Kennan and Walker (2011), there is substantial heterogeneity in moving costs.

In contrast, the SPACE model can match the main facts about internal migration without any moving costs.[30] In other words, the fact that most people do not move is not sufficient evidence to conclude that moving costs are large.

A common counterfactual in the literature is to consider changes in moving costs (Kennan and Walker, 2011; Schubert, 2021; Zerecero, 2021), which is also an actual policy used by some localities.[31] For example, Kennan and Walker (2011) finds that a moving subsidy could substantially increase the gross migration rate. In a moving cost model, a temporary incentive to move to location i has a very persistent impact on the population of i. In contrast, in our model, a moving subsidy would encourage people to relocate, but

[29]Most papers focusing on the United States estimate moving costs of a similar magnitude (Monras, 2018; Bartik and Rinz, 2018). Tombe and Zhu (2019) also estimates very large moving costs in China, on the order of 50 percent of annual income within province, and 90 percent of annual income across provinces. However, their measure of migration is a flow cost, borne by the person every year. Hence, in distinguishing between moving costs and persistent match-specific utilities, it maps more naturally onto a persistent utility to not live outside the origin location.

[30]Including birthplace as a state variable in a moving cost model, which introduces some persistence in match-specific utilities, lowers estimates of moving costs by about 10 percent (Zerecero, 2021).

[31]A handful of cities around the United States offer monetary incentives to relocate (Cornerstone Home Lending, 2021).

only for as long as the subsidy lasts. After the subsidy expires, they are no more likely to remain in the place they moved to, than they would be to live there had the subsidy never occurred.[32]

One particular way of lowering "moving costs" may be improving infrastructure such as roads, which increases migration (Morten and Oliveira, 2018). The SPACE model could be modified to assume that roads increase the correlation of the idiosynchratic shocks between two places, which raises the migration between them, even though moving costs have not gone down. In the moving cost model, the increased migration must reflect lower moving costs, and so welfare would have increased, net the cost of the roads. In the modified SPACE model, though, the welfare effects are much more muted.[33]

4.3 Macro Population Elasticities

Another common use for migration models is to calculate population elasticities to changes in a location's utility, v_i, both in the short-run or the long-run.

In the standard moving cost model, the elasticity of population with respect to v_j is given by:

$$\frac{\partial \log p_i}{\partial v_j} = -\sum_k \frac{m_{i\to k}}{p_i}\frac{m_{k\to j}}{p_j}$$

This is approximately proportional to the migration rate between i and j, since the two terms where $k = i$ or $k = j$ are much larger than the remaining terms.

The elasticities in the two models are similar in that they are approximately proportional to the migration rate when migration rates are low. Given that the scale of v_j is not specified in either model, the constant terms are ignorable without loss of generality.

However, in the long-run, the similarities of population elasticities between the SPACE model and the moving cost model break down. Consider a one-time permanent change in v_{it} for the SPACE model or the moving cost model. In the SPACE model, the population elasticity is still exactly the same, since it is given by corollary 1.

This is not the case with the moving cost model:

Proposition 4. *In steady-state, the long-run elasticity* $\frac{\partial \log p_i}{\partial v_j}$ *of a moving cost model is*

[32] Of course, one could imagine other economic reasons that populations remain higher after a population expansion, such as the accumulation of housing capital or agglomeration in that place.

[33] Of course, other welfare benefits, such as those that come through increased trade, do not depend on how migration is modeled.

not the same as in the short-run. Rather, as $\Delta \to \infty$,

$$\lim_{\Delta \to \infty} \frac{\partial \log p_i}{\partial v_j} = -2p_j \tag{10}$$

when $i \neq j$.

Recall that the total population is mass 1, so p_i is both the population of i and its population share.

Interestingly, these steady-state elasticities are the same as a static logit, and a key difference from the SPACE model is that they have no relationship to migration data. In Appendix B.3, we show equation (10) is a good numerical approximation to the long-run of a calibrated moving cost model, when moving costs lead to realistic migration rates, rather than being infinite.

In the long-run, the moving cost model has no notion that closer states are better substitutes or that states with higher migration are likely to be more impacted by a change in the other state. The moving cost model would not predict that a state with a high migration rate has a more long-run elastic population in response to a policy change than a state with a low migration rate. Rather, the only thing that you need to know is the population share of the state receiving the shock to calculate all the relevant elasticities (approximately).[34]

Figure 6 summarizes the conclusions of this section. In the short-run (Panel a), the SPACE model and the moving cost model predict practically the same cross-elasticities of population. But the in the long-run (Panel b), there is almost no relationship between the two.

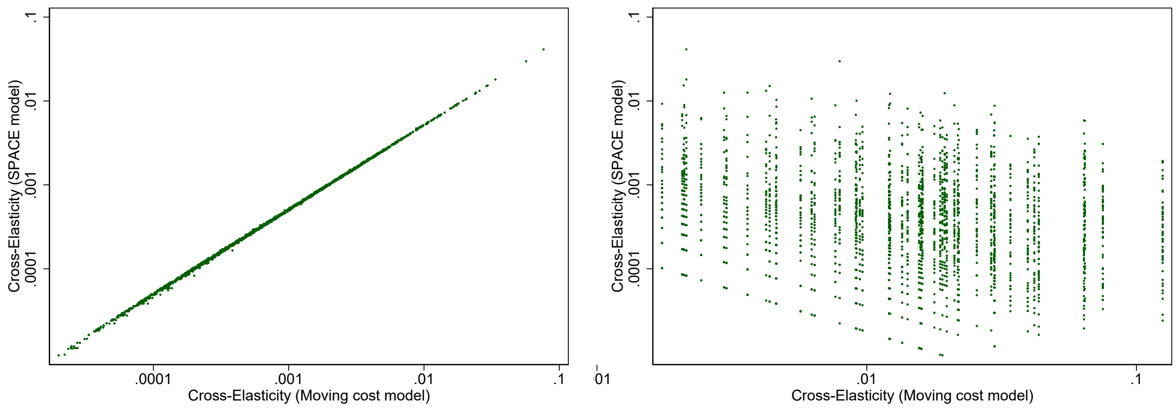
4.4 Macro dynamics

Given the differences in long-run population elasticities, it follows that the intermediate dynamics must also be different across the two models. We illustrate by considering a one-time permanent shock to Louisiana utility, $v_{\text{Louisiana}}$, to see how populations respond over time.

[34]In the moving cost model, the migration rates govern the speed of adjustment (Kleinman et al., 2023), but not the long-run effects.

Including birthplace as a state variable in the moving cost model would mean that adjustments would depend on the population shares of people born in different places. To this extent, migration between states would be correlated to the population cross-elasticities (Zabek, 2020).

Other features of the model can also change the population elasticities. For example, Monte, Redding and Rossi-Hansberg (2018) adds commuting to a static model of location choice to generate variation in population elasticities.



(a) Short-run

(b) Long-run

Figure 6: A comparison of the population cross-elasticities between the SPACE and moving cost models. For both figures, each dot represents a pair of states. The point is located at the population cross-elasticity between the two states in each of the two models. The constant multiplicative terms are ignored, since each model is subject to a normalization of utility. All four axes have log scales.

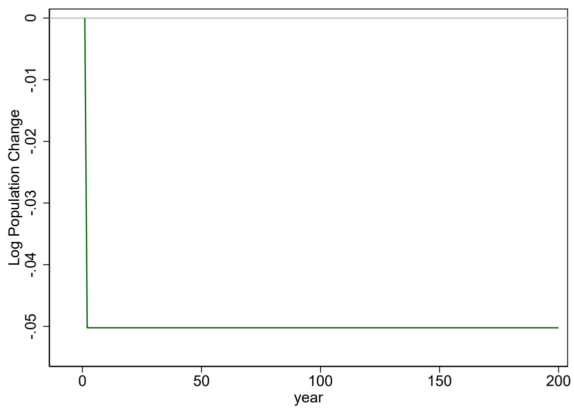
In Figure 7(a) and (b), we show that the population dynamics after a one-time permanent shock are starkly different. In both, we consider a one-time permanent change to the baseline utility of Louisiana, leaving all other states' utilities constant, and we simulate both models for many periods.[35]

In the SPACE model (Panel a), the population adjustment in Louisiana is immediate, and the population stops adjusting after the first period. In contrast, in the moving cost model (Panel b), the population adjustment takes many years, with the model finally getting close a steady-state after almost 200 years.

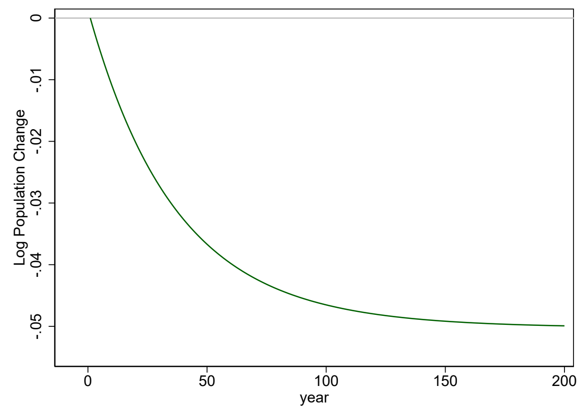
In Panels (c) and (d) we illustrate the dynamics for other states in response to the same shock to Louisiana's utility. In the SPACE model (Panel c), there is a bigger population effect on Mississippi than there is on New York, as one would expect due to the geography. Again the dynamics are immediate. But in Panel (d), the dynamics follow interesting and perhaps unintuitive patterns. In New York, the population adjustment is particularly slow because of low migration between Louisiana and New York. In contrast, for Mississippi, the population dramatically overshoots its long-run steady state because there is so much migration between Louisiana and Mississippi.

This exercise is not necessarily helpful for distinguishing between the two models, but

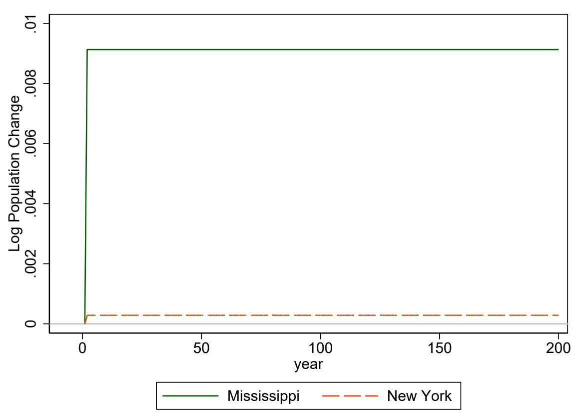
[35] The size of the shocks in each model is normalized to have a long-term effect of about 5 percent of the population for Louisiana. However, the scales are not particularly important, as the focus of this exercise is the dynamics.



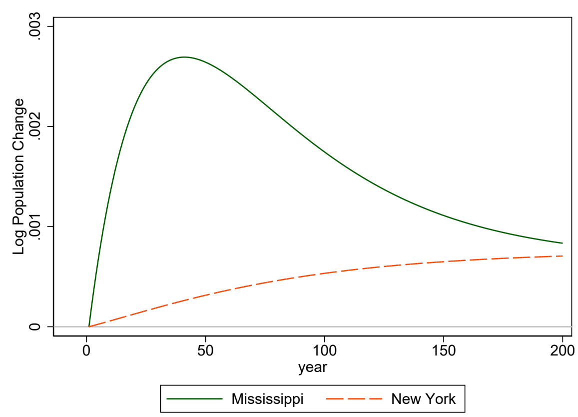
(a) Louisiana, SPACE model



(b) Louisiana, moving cost model



(c) Mississippi and New York, SPACE model



(d) Mississippi and New York, moving cost model

Figure 7: Population Dynamics after a one-time permanent change in $v_{\text{Louisiana}}$, in the SPACE model and the moving cost model, for Louisiana, Mississippi, and New York. Mississippi and New York were chosen to represent two states for which there is high gross migration with Louisiana, and low gross migration with Louisiana, respectively.

it does mean that the models will interpret empirical facts differently. We discuss this more in Section 5.

4.5 Implied Utility Changes

Given the differences in population elasticities, it must be the case that the models will imply different things about changes in utility over time. This is important for papers that wish to estimate the welfare effect of some policy or event that varies across space. For example, Diamond (2016) asks what determines why people live in different places over time.

We can represent the SPACE model's population changes using exact hat algebra. Under our previous calibration, it is given by:

$$\hat{p}_i = \hat{v}_i^{1-\tilde{\rho}} \sum_{j \neq i} \frac{\tilde{w}_{ij}}{p_i} \frac{\tilde{w}_{ij} + \tilde{w}_{ji}}{\tilde{w}_{ij}\hat{v}_i^{1-\tilde{\rho}} + \tilde{w}_{ji}\hat{v}_j^{1-\tilde{\rho}}}$$

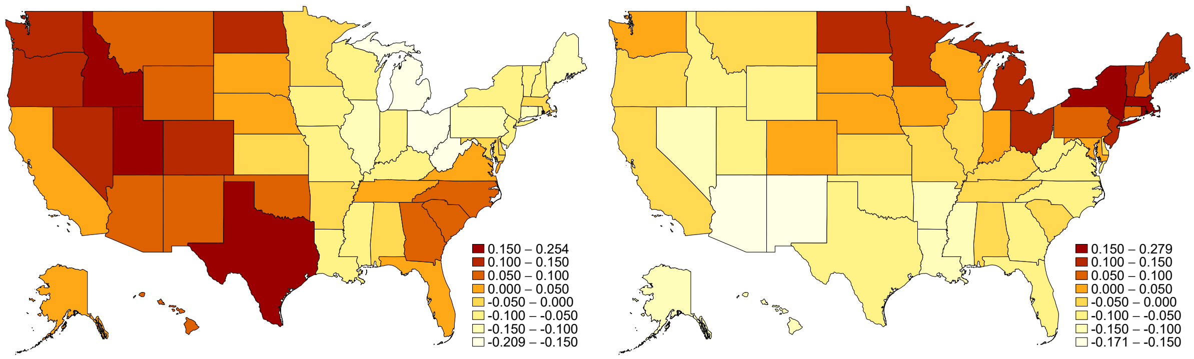
This is straightforward to invert numerically and calculate the $\hat{v}_i$ based on the $\hat{p}_i$. Note that while this exact formula depends on our calibration, it will give the same answer to first-order of any choice of G, since the population elasticities to v are pinned down by migration rates.

In contrast, for the moving cost model, the exact hat algebra is different:

$$\hat{p}_{it} = \hat{v}_i \sum_j \frac{\frac{m_{ij}}{p_i}\hat{p}_{jt-1}}{\sum_k \frac{m_{jk}}{p_j}\hat{v}_k}$$

This formula is similar to the dynamic hat algebra derived by Caliendo et al. (2019). Note that it depends on $\hat{p}_{it-1}$ as well as $\hat{p}_{it}$. For this exact hat algebra, we need to know about the change in how populations are changing, rather than simply the change in population. Again, this is straightforward to invert numerically, given data on $\hat{p}_{it}$ and $\hat{p}_{it-1}$.

To illustrate this, we consider the utility changes implied by the SPACE model and moving cost model from 2005-2018, the span of our data. We show the results in Figure 8. In the SPACE model, the places that have the biggest increase in relative utility are in the South and West, places that have seen large growth in population. New England and the Rust Belt have some of the largest relative decreases. In the moving cost model, the utility changes are almost the opposite. New York and New England have increased in relative utility, while the South and the West have mostly had relative declines. Overall, there is a -0.49 correlation between the log of utility changes implied by the two models.



(a) SPACE model (b) Moving cost model

Figure 8: Change in utilities v_j, 2005-2018, implied by the SPACE model and the moving cost model.

This has important implications for estimating spatial models. For example, if one wanted to estimate the effects of a wage, rent, or amenity change on utility, you would get very different answers using the implied utilities from the SPACE model versus a moving cost model.

5 Discussion

5.1 Relation to Literature

Given the major differences between the two models on several of these questions, it is worth emphasizing why these differences matter. For the micro forecasting and for the interpretation, we think the reasons to care about differences are obvious, but for the population elasticities and the dynamics, it is important to consider the context of the literature.

One big question in the literature is to what extent does population adjust to shocks? For example, if one particular location has a shock that permanently increases the utility of living there, how will that affect the distribution of the population around the country? This is a question that is asked by Caliendo et al. (2019) with respect to the China shock of Autor, Dorn and Hanson (2013), by Giannone (2017) with respect to skill-biased technological change, and by Cruz and Rossi-Hansberg (2021), Oliveira and Pereda (2020), and Rudik et al. (2021) with respect to climate change.[36] Based on the previous

[36] A reader may wonder why we do not replicate one of these papers to highlight the differences. However, doing such a replication would erroneously indicate that the SPACE model is less good at hitting the medium-run dynamics of population adjustment. The reason for this is that even if these

propositions, both models agree that places with high gross migration, such as D.C., have very elastic population to local shocks in the short-run, compared to places with little migration. However, in the long-run, the SPACE model continues to make this prediction, while the moving cost model predicts similar elasticities for all locations. Similarly, in the short-run, both models agree that the population effects are felt in states that have lots of migration between them and the state with the shock. A shock to D.C. will affect Maryland and Virginia more than it will affect Arizona. This is consistent with the "donut" phenomenon during the recent COVID-19 crisis, as areas around major cities have experienced population and house price growth in recent years (Ramani and Bloom, 2021). Again, this holds in the long-run too for the SPACE model, but migration would not generate a long-run donut phenomenon in the moving cost model.

Another key question in this literature is how quickly the migration adjustment takes place (Kleinman et al., 2023; Amior and Manning, 2018; Caliendo et al., 2019). The SPACE model answers this question in that population adjustment occurs as quickly as the baseline utility of a place changes.

In the data, migration is usually quite persistent. For example, the Rust Belt has had low inmigration for decades, and the Sun Belt has had high inmigration for decades. In the moving cost model, much of this persistence is due to the fact that migration is inherently persistent (Kleinman et al., 2023), i.e. the Rust Belt had a large negative utility shock a long time ago, and the process of moving out has been very slow.

The SPACE model interprets this fact as being about the utility of a location adjusting slowly. It could be that the underlying shocks to utility are slow. Or there was still a big initial shock, but some equilibrium force makes utility fall slowly. For example, housing is durable, and so housing becomes cheap as people move out, keeping utility from falling too quickly (Glaeser and Gyourko, 2005). Similarly, there may be similar mechanisms through the labor market that make structural transformation slow.[37] Or it could be that

models are not explicitly targeting the medium-run dynamics, they do get to choose what features of the world to add and can choose to include or not include features that will get the dynamics right. For example, Glaeser and Gyourko (2005) argues that the reason declining cities decline slowly is because the housing stock is slow to depreciate. Many of the quantitative papers do not have this feature. So of course, subbing out the migration block from those models would lead to unrealistic dynamics, if we did not also add in a feature like the one in Glaeser and Gyourko (2005). Sometimes, people tell us that they think of the moving costs as representing these other features, such as housing depreciation or labor market frictions. In that case, we think it is better to explicitly model them. However, we do not think such a model is within the scope of this paper. Writing a completely new model, while it might add to the point about macro elasticities, would distract from highlighting the important differences in micro forecasting and interpretation.

[37]Kleinman et al. (2023) discuss how having location specific durable capital can keep wages high after a negative productivity shock.

initial shocks are small, but then as people move in, the effects on utility become amplified with a delay (Howard, 2020). The SPACE model would emphasize these various forces as reasons for persistence in migration, whereas the moving cost model would attribute the persistence to an inherent property of migration itself, and find less explanatory power for these forces.

Other papers are concerned with the effects of location-specific shocks on aggregate outcomes such as welfare or output (Tombe and Zhu, 2019; Eckert and Peters, 2018; Hsieh and Moretti, 2019). The degree to which people are able to move is an important factor for these outcomes. Because the second order effects are determined by these population elasticities, it shows that a shock that affects a higher-gross-migration place will have larger total effects on welfare if the shock is positive, and smaller effects if the shock is negative. Again, this holds in the short-run for both models, but in the long-run only for the SPACE model.

Finally, the spatial correlation of a shock is an important determinant of its welfare consequences. If a negative shock is extremely localized, it may be easy to move away from it, and there will be lots of insurance. If shocks are correlated across space, then the welfare effects may be much less insurable. Of course, in the long-run of a moving cost model, this effect will no longer hold.

5.2 More complex moving cost models

In Section 4, we compared the SPACE model to a very simple version of a moving cost model. Yet as mentioned in the literature review, there is significant research that enriches the moving cost model to match a variety of facts (Kaplan and Schulhofer-Wohl, 2017; Giannone et al., 2020; Porcher, 2020; Mangum and Coate, 2019; Zerecero, 2021; Monras, 2018). Kennan and Walker (2011) includes features to increase home bias and return migration.

As we showed in Section 2.1, return migration, home bias, nor age are sufficient features to hit the square root fact. So the puzzle that motivated the model is not dependent on us having considered a simple version of the moving cost model. But what about evaluating the differences between the moving cost model and the SPACE model? Are the conclusions that the choice of model is important robust to considering richer versions of the moving cost model? Here, we argue that the answer is yes. In the rest of this section, we consider each of the differences we highlighted before.

For the prediction of individuals' locations, the reason that the SPACE model outperformed the moving cost model was because it could hit the square root fact. So if extensions of the moving cost model still do not hit the square root fact, they might be an improvement at predicting locations, but are not going to make the same predictions as the SPACE model.

For the interpretation of why people rarely move, some of the additional features imply lower estimated moving costs (Zerecero, 2021; Giannone et al., 2020), but never by orders of magnitude. So the difference between the SPACE model—which has no moving costs—and any moving cost model will remain large.

For population elasticities, more complex moving cost models feature long-run elasticities which may not necessarily be the same as a static logit model. For example, models with home bias have more similar elasticities to the SPACE model than the baseline moving cost model does.[38] However, except by coincidence, none of the additional features would generate the feature that the long-run and short-run population elasticities are the same. So the dramatic difference between the SPACE model and the moving cost model will remain, both for long-run elasticities and for dynamics.

For the implied utility changes, the exact implied utilities will obviously change with a richer model. However, it doesn't change the fundamental fact that the implicit utility is a function of migration in the moving cost model, but populations in the SPACE model.

6 Conclusion

We use a dataset of credit reports to document a new fact: the t-year migration rate is proportional to the square root of t. We propose a new model to match this fact, which has different implications for many economic questions than the standard moving cost model.

[38]With home bias, the elasticities are given by $\lim_{\Delta\to\infty} \partial \log p_j / \partial v_k = -\sum_i w_{ij} p_{ik}$, where $w_{ij} = \frac{p_{ij}}{p_j}$ is the share of people in j who are from i, and p_{ik} is the share of people from i living in j. These population shares are likely correlated to the amount of bilateral migration. However, it is a different formula, and the population share levels are likely different than the migration rates, so the dynamics in the two models will still be different.

References

Allen, Treb and Dave Donaldson, "Persistence and path dependence in the spatial economy," 2020. National Bureau of Economic Research Working Paper.

Amior, Michael and Alan Manning, "The persistence of local joblessness," *American Economic Review*, 2018, *108* (7), 1942–70.

Autor, David, David Dorn, and Gordon H Hanson, "The China syndrome: Local labor market effects of import competition in the United States," *American Economic Review*, 2013, *103* (6), 2121–68.

Bartik, Alexander W and Kevin Rinz, "Moving costs and worker adjustment to changes in labor demand: Evidence from longitudinal census data," 2018.

Bayer, Christian and Falko Juessen, "On the dynamics of interstate migration: Migration costs and self-selection," *Review of Economic Dynamics*, 2012, *15* (3), 377–401.

Blanchard, Olivier Jean and Lawrence F Katz, "Regional evolutions," *Brookings Papers on Economic Activity*, 1992, *1992* (1), 1–75.

Bryan, Gharad and Melanie Morten, "The aggregate productivity effects of internal migration: Evidence from Indonesia," *Journal of Political Economy*, 2019, *127* (5), 2229–2268.

Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro, "Trade and labor market dynamics: General equilibrium analysis of the china trade shock," *Econometrica*, 2019, *87* (3), 741–835.

Census, U.S., "Income, Poverty and Health Insurance Coverage in the United States: 2010," 2011.

Coen-Pirani, Daniele, "Understanding gross worker flows across US states," *Journal of Monetary Economics*, 2010, *57* (7), 769–784.

Correia, Sergio, Paulo Guimaraes, and Thomas Zylkin, "PPMLHDFE: Stata module for Poisson pseudo-likelihood regression with multiple levels of fixed effects," 2019.

Cruz, José-Luis and Esteban Rossi-Hansberg, "The economic geography of global warming," *University of Chicago, Becker Friedman Institute for Economics Working Paper*, 2021, (2021-130).

Davis, Morris A, Jonas DM Fisher, and Marcelo Veracierto, "Migration and urban economic dynamics," *Journal of Economic Dynamics and Control*, 2021, *133*, 104234.

DeWaard, Jack, Janna Johnson, and Stephan Whitaker, "Internal migration in the United States: A comprehensive comparative assessment of the Consumer Credit Panel," *Demographic research*, 2019, *41*, 953.

_ , Mathew Hauer, Elizabeth Fussell, Katherine J Curtis, Stephan Whitaker, Kathryn McConnell, Kobie Price, and David Egan-Robertson, "User Beware: Concerning Findings from Recent US Internal Revenue Service Migration Data," 2020. Minnesota Population Center Working Paper No. 2020-02.

Diamond, Rebecca, "The determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000," *American Economic Review*, 2016, *106* (3), 479–524.

Eckert, Fabian and Michael Peters, “Spatial Structural Change,” 2018.
Farrokhi, Farid and David Jinkins, “Dynamic amenities and path dependence in location choice: Evidence from danish refugee placement,” 2021.
Fonseca, Julia, “Less Mainstream Credit, More Payday Borrowing? Evidence from Debt Collection Restrictions,” *Journal of Finance*, 2022.
— **and Jialan Wang**, “How Much do Small Businesses Rely on Personal Credit?,” 2022.
Frobenius, Ferdinand Georg, “Über Matrizen aus nicht negativen Elementen,” *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften*, 1912, pp. 456–477.
Fudenberg, Drew, Wayne Gao, and Annie Liang, “How flexible is that functional form? Quantifying the restrictiveness of theories,” 2023.
Fujiwara, Thomas, Eduardo Morales, and Charly Porcher, “A Revealed Preference Approach to Measuring Information Frictions in Migration Decisions,” 2022.
Giannone, Elisa, “Skill-Biased Technical Change and Regional Convergence,” 2017.
— **, Qi Li, Nuno Paixao, and Xinle Pang**, “Unpacking Moving,” 2020.
Gies Consumer and Small Business Credit Panel, 2004-2018. Accessed July 20, 2021.
Glaeser, Edward L and Joseph Gyourko, “Urban decline and durable housing,” *Journal of Political Economy*, 2005, *113* (2), 345–375.
Han, Peter, “The Bank Premium and Market Concentration in the Mortgage Market,” 2022.
Hao, Tongtong, Ruiqi Sun, Trevor Tombe, and Xiaodong Zhu, “The effect of migration policy on growth, structural change, and regional inequality in China,” *Journal of Monetary Economics*, 2020, *113*, 112–134.
Heise, Sebastian and Tommaso Porzio, “The aggregate and distributional effects of spatial frictions,” 2021. National Bureau of Economic Research Working Paper.
Howard, Greg, “The migration accelerator: Labor mobility, housing, and demand,” *American Economic Journal: Macroeconomics*, 2020, *12* (4), 147–79.
— **and Jack Liebersohn**, “Why is the rent so darn high? The role of growing demand to live in housing-supply-inelastic cities,” *Journal of Urban Economics*, 2021, *124*, 103369.
— **,** — **, and Adam Ozimek**, “The Short-and Long-Run Effects of Remote Work on US Housing Markets,” *Journal of Financial Economics*, 2023.
Hsieh, Chang-Tai and Enrico Moretti, “Housing constraints and spatial misallocation,” *American Economic Journal: Macroeconomics*, 2019, *11* (2), 1–39.
IRS Migration Data, “Internal Revenue Service Statistics of Income Tax Statistics - Migration Data,” 2004-2018. https://www.irs.gov/statistics/soi-tax-stats-migration-data.
Jia, Ning, Raven Molloy, Christopher Smith, and Abigail Wozniak, “The economics of internal migration: Advances and policy questions,” *Journal of Economic Literature*, 2023, *61* (1), 144–180.
Kaplan, Greg and Sam Schulhofer-Wohl, “Understanding the long-run decline in interstate migration,” *International Economic Review*, 2017, *58* (1), 57–94.
Kennan, John and James R Walker, “The effect of expected income on individual migration decisions,” *Econometrica*, 2011, *79* (1), 211–251.

Kleemans, Marieke, "Migration choice under risk and liquidity constraints," 2015.
Kleinman, Benny, Ernest Liu, and Stephen Redding, "Dynamic spatial general equilibrium," *Econometrica*, 2023.
Koşar, Gizem, Tyler Ransom, and Wilbert Van der Klaauw, "Understanding migration aversion using elicited counterfactual choice probabilities," *Journal of Econometrics*, 2021.
Lending, Inc. Cornerstone Home, "These 11 U.S. Cities Will Pay You to Live There," 2021. Accessed August 9, 2022. https://www.houseloanblog.net/get-paid-to-relocate/.
Lind, Nelson and Natalia Ramondo, "Trade with Correlation," *American Economic Review*, February 2023, *113* (2), 317–53.
Mangum, Kyle and Patrick Coate, "Fast locations and slowing labor mobility," 2019.
McFadden, Daniel, "Modelling the choice of residential location," 1977.
Molloy, Raven, Christopher L Smith, and Abigail Wozniak, "Internal migration in the United States," *Journal of Economic Perspectives*, 2011, *25* (3), 173–96.
Monras, Joan, "Economic shocks and internal migration," 2018.
Monte, Ferdinando, Stephen J Redding, and Esteban Rossi-Hansberg, "Commuting, migration, and local employment elasticities," *American Economic Review*, 2018, *108* (12), 3855–90.
Morris-Levenson, Joshua and Marta Prato, "The Origins of Regional Specialization." PhD dissertation, The University of Chicago 2022.
Morten, Melanie and Jaqueline Oliveira, "The effects of roads on trade and migration: Evidence from a planned capital city," *NBER Working Paper*, 2018, *22158*, 1–64.
Oliveira, Jaqueline and Paula Pereda, "The impact of climate change on internal migration in Brazil," *Journal of Environmental Economics and Management*, 2020, *103*, 102340.
Panel Survey of Income Dynamics, 1969-1997. https://psidonline.isr.umich.edu/.
Perron, Oskar, "Zur Theorie der Matrizen," *Mathematische Annalen*, 1907, *64* (2), 248–263.
Porcher, Charly, "Migration with costly information," 2020.
Ramani, Arjun and Nicholas Bloom, "The Donut effect of COVID-19 on cities," 2021. National Bureau of Economic Research Working Paper.
Roback, Jennifer, "Wages, rents, and the quality of life," *Journal of Political Economy*, 1982, *90* (6), 1257–1278.
Rogerson, Peter A, "A new method for finding geographic centers, with application to US States," *The Professional Geographer*, 2015, *67* (4), 686–694.
Rosen, Sherwin, "Wage-based indexes of urban quality of life," *Current Issues in Urban Economics*, 1979, pp. 74–104.
Rudik, Ivan, Gary Lyn, Weiliang Tan, and Ariel Ortiz-Bobea, "The Economic Effects of Climate Change in Dynamic Spatial Equilibrium," 2021.
Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek, "Integrated Public Use Microdata Series: Version 6.0 [dataset]," `http://doi.org/10.18128/D010.V6.0` 2015. Minneapolis: University of Minnesota.

Saks, Raven E and Abigail Wozniak, "Labor reallocation over the business cycle: New evidence from internal migration," *Journal of Labor Economics*, 2011, *29* (4), 697–739.

Schubert, Gregor, "House Price Contagion and US City Migration Networks," 2021.

Silva, JMC Santos and Silvana Tenreyro, "The log of gravity," *The Review of Economics and Statistics*, 2006, *88* (4), 641–658.

Tombe, Trevor and Xiaodong Zhu, "Trade, migration, and productivity: A quantitative analysis of china," *American Economic Review*, 2019, *109* (5), 1843–72.

Vincenty, Thaddeus, "Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations," *Survey review*, 1975, *23* (176), 88–93.

Zabek, Michael A, "Local ties in spatial equilibrium," 2020.

Zerecero, M, "The Birthplace Premium," 2021.

Online Appendix

A Proofs

A.1 Proof of Proposition 1

Proof: In the steady-state of the moving cost model, consider the probability of someone living in i living in j after t years. This probability is given by:

$$\frac{m_{i\to j}^t}{p_i} = \sum_{k_1,k_2,k_3,\ldots,k_{t-1}\in I} \frac{m_{i\to k_1}}{p_i}\left(\prod_{s=1}^{t-2}\frac{m_{k_s\to k_{s+1}}}{p_{k_s}}\right)\frac{m_{k_{t-1}\to j}}{p_{k_{t-1}}}$$

Consider

$$\lim_{\Delta\to\infty}\frac{m_{i\to j}^t}{m_{i\to j}} = \lim_{\Delta\to\infty}\sum_{k_1,k_2,k_3,\ldots,k_{t-1}\in I}\frac{\frac{m_{i\to k_1}}{p_i}\left(\prod_{s=1}^{t-2}\frac{m_{k_s\to k_{s+1}}}{p_{k_s}}\right)\frac{m_{k_{t-1}\to j}}{p_{k_{t-1}}}}{\frac{m_{i\to j}}{p_i}}$$

We can calculate each term in the summation. First, consider all the summations such that there exists a cutoff T such that for all $s < T$, $k_s = i$ and for all $s \geq T$, $k_s = j$. There are t such combinations. For each combination, the product in the numerator is a lot of migration probabilities from i to i (non-migration), and one migration probability from i to j. For the non-migration probabilities, the limit as $\delta \to \infty$ of $\frac{m_{i\to i}}{p_i}$ or $\frac{m_{j\to j}}{p_j}$ is 1. The remaining term, $\frac{m_{i\to j}}{p_i}$ cancels with the denominator, so the limit is 1.

Next consider all other terms. For each, there is one year in which the person moves away from i and another year in which they move from their next location to somewhere else. Put these two terms first in the product:

$$\frac{m_{i\to\ell}/p_i}{m_{i\to j}/p_i}\cdot\frac{m_{\ell\to k_s}}{p_\ell}\cdot\ldots$$

The first fraction is $\frac{\exp(v_{\ell t}-\delta_{i\ell})}{\exp(v_{jt}-\delta_{ij})}$ regardless of Δ, which is a constant. The second term converges to zero when $\Delta\to\infty$. All the following terms are between 1 and 0, so the whole product converges to 0.

Therefore, the sum is over t 1's and many zeros. Hence,

$$\lim_{\Delta\to\infty}\frac{m_{i\to j,t}}{m_{i\to j}} = t$$

□

A.2 Proof of Proposition 2

We wish to show that:

$$m_{i\to j} = (1-\rho)e^{v_i+v_j}\left(\frac{G_i(e^{v_1},...,e^{v_I})G_j(e^{v_1},...,e^{v_I})}{G(e^{v_1},...,e^{v_I})^2} - \frac{G_{ij}(e^{v_1},...,e^{v_I})}{G(e^{v_1},...,e^{v_I})}\right)$$

The definition of migration is

$$P(v_i + \epsilon_{it} \geq \max_k v_k + \epsilon_{it}, v_j + \epsilon_{jt+1} \geq \max_k v_k + \epsilon_{kt+1})$$

when $(\epsilon, \epsilon_{t+1}) \sim F_2$. We can express this as an integral:

$$\int_{\substack{v_i+\epsilon_{it}\geq \max_k v_k+\epsilon_{kt} \\ v_j+\epsilon_{jt+1}\geq \max_k v_k+\epsilon_{kt}}} dF_2(\epsilon_t, \epsilon_{t+1})$$

We can integrate out all the ϵ's except ϵ_{it} and ϵ_{jt+1}, which leaves us with just the cross-partial derivative of F_2 with respect to the remaining two ϵ's, evaluated at $v_i + \epsilon_{it} - v_k$ for all the ϵ_{kt} or $v_j + \epsilon_{jt+1} - v_k$ for all the ϵ_{kt+1}:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\partial^2 F_2}{\partial \epsilon_{it}\partial \epsilon_{jt+1}}(v_i - v_1 + \epsilon_{it}, ..., v_I - v_1 + \epsilon_{it}, v_j - v_1 + \epsilon_{It+1}, ..., v_j - v_I + \epsilon_{jt+1})d\epsilon_{it}d\epsilon_{jt+1}$$

Recall that

$$F_2 = \exp(-G((e^{-\epsilon_{1t}/(1-\rho)} + e^{-\epsilon_{1t+1}/(1-\rho)})^{1-\rho}, ..., (e^{-\epsilon_{It}/(1-\rho)} + e^{-\epsilon_{It+1}/(1-\rho)})^{1-\rho})))$$

So

$$\begin{aligned}\frac{\partial^2 F_2}{\partial \epsilon_{it}\partial \epsilon_{jt+1}} =& \exp(-G)G_iG_j(e^{-\epsilon_{it}/(1-\rho)} + e^{-\epsilon_{it+1}/(1-\rho)})^{-\rho}e^{-\epsilon_{it}/(1-\rho)}(e^{-\epsilon_{jt}/(1-\rho)} + e^{-\epsilon_{jt+1}/(1-\rho)})^{-\rho}e^{-\epsilon_{jt+1}/(1-\rho)} \\ &+ \exp(-G)G_{ij}(e^{-\epsilon_{it}/(1-\rho)} + e^{-\epsilon_{it+1}/(1-\rho)})^{-\rho}e^{-\epsilon_{it}/(1-\rho)}(e^{-\epsilon_{jt}/(1-\rho)} + e^{-\epsilon_{jt+1}/(1-\rho)})^{-\rho}e^{-\epsilon_{jt+1}/(1-\rho)}\end{aligned}$$

G is homothetic of degree 1, so G_i is homothetic of degree 0 and G_{ij} is homothetic of degree -1. So

$$\begin{aligned}&G((e^{-(v_i-v_1+\epsilon_{it})/(1-\rho)} + e^{-(v_j-v_1+\epsilon_{jt+1})/(1-\rho)})^{1-\rho}, ..., (e^{-(v_i-v_I+\epsilon_{it})/(1-\rho)} + e^{-(v_j-v_I+\epsilon_{jt+1})/(1-\rho)})^{1-\rho}) \\ &\qquad = G(e^{v_1}, ..., e^{v_I})(e^{-(v_i+\epsilon_{it})/(1-\rho)} + e^{-(v_j+\epsilon_{jt+1})/(1-\rho)})^{1-\rho}\end{aligned}$$

and

$$G_i((e^{-(v_i-v_1+\epsilon_{it})/(1-\rho)}+e^{-(v_j-v_1+\epsilon_{jt+1})/(1-\rho)})^{1-\rho},...,(e^{-(v_i-v_I+\epsilon_{it})/(1-\rho)}+e^{-(v_j-v_I+\epsilon_{jt+1})/(1-\rho)})^{1-\rho})$$
$$=G_i(e^{-v_1},...,e^{-v_I})$$

and

$$G_{ij}((e^{-(v_i-v_1+\epsilon_{it})/(1-\rho)}+e^{-(v_j-v_1+\epsilon_{jt+1})/(1-\rho)})^{1-\rho},...,(e^{-(v_i-v_I+\epsilon_{it})/(1-\rho)}+e^{-(v_j-v_I+\epsilon_{jt+1})/(1-\rho)})^{1-\rho})$$
$$=\frac{G_{ij}(e^{v_1},...,e^{v_I})}{(e^{-(v_i+\epsilon_{it})/(1-\rho)}+e^{-(v_j+\epsilon_{jt+1})/(1-\rho)})^{1-\rho}}$$

Since $G(v_1,...,v_I)$ is just a constant, let us denote it by $\bar{G}$. Similarly, denote $\bar{G}_i \equiv G_i(v_1,...,v_I)$ and $\bar{G}_{ij} \equiv G_{ij}(v_1,...,v_I)$.

So our integral becomes

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} & \exp(-\bar{G}(e^{-(v_i+\epsilon_{it})/(1-\rho)}+e^{-(v_j+\epsilon_{jt+1})/(1-\rho)})^{1-\rho})\bar{G}_i\bar{G}_j \\ & \times (e^{-\epsilon_{it}/(1-\rho)}+e^{-(\epsilon_{jt+1}+v_j-v_i)/(1-\rho)})^{-\rho}e^{-\epsilon_{it}/(1-\rho)}(e^{-(\epsilon_{it}+v_i-v_j)/(1-\rho)}+e^{-\epsilon_{jt+1}/(1-\rho)})^{-\rho}e^{-\epsilon_{jt+1}/(1-\rho)} \\ & + \exp(-\bar{G}(e^{-(v_i+\epsilon_{it})/(1-\rho)}+e^{-(v_j+\epsilon_{jt+1})/(1-\rho)})^{1-\rho})\bar{G}_{ij} \\ & \times \frac{1}{(e^{-(v_i+\epsilon_{it})/(1-\rho)}+e^{-(v_j+\epsilon_{jt+1})/(1-\rho)})^{1-\rho}} \\ & \times (e^{-\epsilon_{it}/(1-\rho)}+e^{-\epsilon_{it+1}/(1-\rho)})^{-\rho}e^{-\epsilon_{it}/(1-\rho)}(e^{-\epsilon_{jt}/(1-\rho)}+e^{-\epsilon_{jt+1}/(1-\rho)})^{-\rho}e^{-\epsilon_{jt+1}/(1-\rho)}d\epsilon_{it}d\epsilon_{jt+1}\end{aligned}$$

This can be simplified:

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} & \exp(-\bar{G}(e^{-(v_i+\epsilon_{it})/(1-\rho)}+e^{-(v_j+\epsilon_{jt+1})/(1-\rho)})^{1-\rho}) \\ & \times \left((\bar{G}_i\bar{G}_j+\bar{G}_{ij}\frac{1}{(e^{-(v_i+\epsilon_{it})/(1-\rho)}+e^{-(v_j+\epsilon_{jt+1})/(1-\rho)})^{1-\rho}}\right) \\ & \times e^{v_i+v_j}(e^{-(\epsilon_{it}+v_i)/(1-\rho)}+e^{-(\epsilon_{jt+1}+v_j)/(1-\rho)})^{-2\rho}e^{-(\epsilon_{it}+v_i)/(1-\rho)}e^{-(\epsilon_{jt+1}+v_j)(1-\rho)}d\epsilon_{it}d\epsilon_{jt+1}\end{aligned}$$

Consider the following u-substitution:

$$\begin{aligned}u &= (e^{-(v_i+\epsilon_{it})/(1-\rho)}+e^{-(v_j+\epsilon_{jt+1})/(1-\rho)})^{1-\rho} \\ du &= (e^{-(v_i+\epsilon_{it})/(1-\rho)}+e^{-(v_j+\epsilon_{jt+1})/(1-\rho)})^{-\rho}e^{-(v_j+\epsilon_{jt+1})/(1-\rho)}d\epsilon_{jt+1}\end{aligned}$$

This leads to a lot of simplification:

$$\int_{-\infty}^{\infty}\int_{e^{-(v_i+\epsilon_{it})}}^{\infty} \exp(-\bar{G}u)\left((\bar{G}_i\bar{G}_j + \bar{G}_{ij}\frac{1}{u}\right)e^{v_i+v_j}u^{-\rho/(1-\rho)}e^{-(\epsilon_{it}+v_i)/(1-\rho)}dud\epsilon_{it}$$

We can change the order of integration:

$$\int_{0}^{\infty}\int_{-v_i-\log u}^{\infty} \exp(-\bar{G}u)\left((\bar{G}_i\bar{G}_j + \bar{G}_{ij}\frac{1}{u}\right)e^{v_i+v_j}u^{-\rho/(1-\rho)}e^{-(\epsilon_{it}+v_i)/(1-\rho)}d\epsilon_{it}du$$

And evaluate the interior integral:

$$\int_{0}^{\infty} \exp(-\bar{G}u)\left((\bar{G}_i\bar{G}_j + \bar{G}_{ij}\frac{1}{u}\right)e^{v_i+v_j}u^{-\rho/(1-\rho)}[(1-\rho)e^{-(\epsilon_{it}+v_i)/(1-\rho)}]_{-v_i-\log u}^{\infty}du$$

Plugging in,

$$(1-\rho)e^{v_i+v_j}\int_{0}^{\infty} \exp(-\bar{G}u)\left((\bar{G}_i\bar{G}_j + \bar{G}_{ij}\frac{1}{u}\right)u^{-\rho/(1-\rho)}u^{1/(1-\rho)}du$$

Simplifying,

$$(1-\rho)e^{v_i+v_j}\int_{0}^{\infty} \exp(-\bar{G}u)\left((\bar{G}_i\bar{G}_j u + \bar{G}_{ij}\right)du$$

Evaluation the integral,

$$(1-\rho)e^{v_i+v_j}\left[-u\frac{\bar{G}_i\bar{G}_j}{\bar{G}}e^{-\bar{G}u} - \frac{\bar{G}_i\bar{G}_j}{\bar{G}^2}e^{-\bar{G}u} - \frac{\bar{G}_{ij}}{\bar{G}}e^{-\bar{G}u}\right]_0^{\infty}$$

This simplifies to:

$$(1-\rho)e^{v_i+v_j}\left(\frac{\bar{G}_i\bar{G}_j}{\bar{G}^2} + \frac{\bar{G}_{ij}}{\bar{G}}\right)$$

To see the other formulation, recall that $p_i = e^{v_i}\frac{\bar{G}_i}{\bar{G}}$. If we factor that out,

$$(1-\rho)p_ip_j\left(1 + \frac{\bar{G}_{ij}\bar{G}}{\bar{G}_i\bar{G}_j}\right)$$

□

A.3 Proof of Corollary 1

The derivative of p_i with respect to v_j is:

$$\frac{\partial p_i}{\partial v_j} = e^{v_i+v_j}\left(-\frac{G_i(e^{v_1},...,e^{v_I})G_j(e^{v_1},...,e^{v_I})}{G(e^{v_1},...,e^{v_I})^2} + \frac{G_{ij}(e^{v_1},...,e^{v_I})}{G(e^{v_1},...,e^{v_I})}\right) \tag{11}$$

By equation (6), we can substitute $\frac{1}{1-\rho}m_{i\to j}$ for the right-hand side. □

A.4 Exact Hat Algebra Derivation

Consider the calibration from Section 3.1. Here, we show the derivation for the exact hat algebra. We would like to express $\hat{p}_i \equiv p_i'/p_i$ as a function of $\hat{v}_i \equiv \exp((v_i' - v_i)/(1-\rho))$ in the limit as $\rho \to 1$.

First, assume that G does not change with v_i as $\rho \to 1$. We can verify this at the end by checking whether $\sum_i p_i\hat{p}_i = 1$. Under this assumption, we only have to consider changes in $G_i e^{v_i}$. This quantity is given by:

$$p_i' = \sum_{j\neq i}\left((w_{ij}e^{v_i'})^{\frac{1}{1-\gamma}} + (w_{ji}e^{v_j'})^{\frac{1}{1-\gamma}}\right)^{-\gamma}(w_{ij}e^{v_i'})^{\frac{1}{1-\gamma}}$$

This can be rewritten

$$p_i' = \sum_{j\neq i}\left((w_{ij}e^{v_i})^{\frac{1}{1-\gamma}}\hat{v}_i^{1-\tilde{\rho}} + (w_{ji}e^{v_j})^{\frac{1}{1-\gamma}}\hat{v}_j^{1-\tilde{\rho}}\right)^{-\gamma}(w_{ij}e^{v_i})^{\frac{1}{1-\gamma}}\hat{v}_i^{1-\tilde{\rho}}$$

Recall that $w_{ij}e^{v_i} = \tilde{w}_{ij}^{1-\gamma}(\tilde{w}_{ij}+\tilde{w}_{ji})^{\gamma}$.

$$p_i' = \sum_{j\neq i}\left(\tilde{w}_{ij}(\tilde{w}_{ij}+\tilde{w}_{ji})^{\frac{\gamma}{1-\gamma}}\hat{v}_i^{1-\tilde{\rho}} + \tilde{w}_{ji}(\tilde{w}_{ij}+\tilde{w}_{ji})^{\frac{\gamma}{1-\gamma}}\hat{v}_j^{1-\tilde{\rho}}\right)^{-\gamma}\tilde{w}_{ij}(\tilde{w}_{ij}+\tilde{w}_{ji})^{\frac{\gamma}{1-\gamma}}\hat{v}_i^{1-\tilde{\rho}}$$

Simplifying,

$$p_i' = \sum_{j\neq i}\left(\tilde{w}_{ij}\hat{v}_i^{1-\tilde{\rho}} + \tilde{w}_{ji}\hat{v}_j^{1-\tilde{\rho}}\right)^{-\gamma}\tilde{w}_{ij}(\tilde{w}_{ij}+\tilde{w}_{ji})^{\gamma}\hat{v}_i^{1-\tilde{\rho}}$$

Recall that as $\rho \to 1$, $\gamma \to 1$ as well, so

$$p_i' = \sum_{j\neq i}\tilde{w}_{ij}\frac{\tilde{w}_{ij}+\tilde{w}_{ji}}{\tilde{w}_{ij}\hat{v}_i^{1-\tilde{\rho}} + \tilde{w}_{ji}\hat{v}_j^{1-\tilde{\rho}}}\hat{v}_i^{1-\tilde{\rho}}$$

To express it as $\hat{p}_i$, divide both sides by p_i:

$$\hat{p}_i = \sum_{j \neq i} \frac{\tilde{w}_{ij}}{p_i} \frac{\tilde{w}_{ij} + \tilde{w}_{ji}}{\tilde{w}_{ij}\hat{v}_i^{1-\tilde{\rho}} + \tilde{w}_{ji}\hat{v}_j^{1-\tilde{\rho}}} \hat{v}_i^{1-\tilde{\rho}}$$

A.5 Proof of Lemma 1

Before we prove Lemma 1, we will need the following lemma.

Lemma 2. *Define Υ to be the cumulative distribution function of the conditional distribution $\Delta\epsilon_n = \frac{\epsilon_{nt+1} - \epsilon_{nt}}{1-\rho}$ given ϵ_{nt}. Then,*

$$\lim_{\rho \to 1} \Upsilon(\Delta\epsilon_1, \Delta\epsilon_2, \Delta\epsilon_3, ...|\epsilon_{1t}, \epsilon_{2t}, \epsilon_{3t}, ...) = \frac{1}{1 + e^{\Delta\epsilon_1}} \frac{1}{1 + e^{\Delta\epsilon_2}} \frac{1}{1 + e^{\Delta\epsilon_3}} ...$$

Proof: The conditional distribution of ϵ_{t+1} given ϵ_t is given by

$$\frac{\frac{\partial^I F_2}{\prod_i \partial\epsilon_{it}}(\epsilon_{1t}, \epsilon_{2t}, ...\epsilon_{It}, \epsilon_{1t+1}, \epsilon_{2t+1}, ...\epsilon_{It+1})}{\frac{\partial^I F_2}{\prod_i \partial\epsilon_{it}}(\epsilon_{1t}, \epsilon_{2t}, ...\epsilon_{It}, \infty, \infty, ..., \infty)}$$

which is

$$\frac{f\left(-(1-\rho)\log(e^{-\frac{\epsilon_{1t}}{1-\rho}} + e^{-\frac{\epsilon_{1t+1}}{1-\rho}}), -(1-\rho)\log(e^{-\frac{\epsilon_{2t}}{1-\rho}} + e^{-\frac{\epsilon_{2t+1}}{1-\rho}})...\right)}{f(\epsilon_{1t}, \epsilon_{2t}, ...)} \prod_i (1 + e^{-\frac{\epsilon_{it+1} - \epsilon_{it}}{1-\rho}})^{-\rho}$$

This means that the CDF of $\Delta\epsilon$ is given by

$$\Upsilon = \frac{f\left(\epsilon_{1t} - (1-\rho)\log(1 + e^{-\Delta\epsilon_1}), \epsilon_{2t} - (1-\rho)\log(1 + e^{-\Delta\epsilon_2}), ...\right)}{f(\epsilon_{1t}, \epsilon_{2t}, ...)} \prod_i (1 + e^{-\Delta\epsilon_i})^{-\rho}$$

Taking the limit as $\rho \to 1$ gives us:

$$\lim_{\rho \to 1} \Upsilon = \prod_i \frac{1}{1 + e^{-\Delta\epsilon_i}}$$

□

This lemma implies that when ρ is close to 1, the changes in the ϵ's are well-approximated by independent logistic distributions.

Now we turn to Lemma 1. Migration is is given by:

$$m_{ij}^s = \mathbb{P}(v_i + \epsilon_{in0} > \max_k v_k + \epsilon_{kn0} \text{ and } v_j + \epsilon_{jns} > \max_k v_k + \epsilon_{kns})$$

$$m_{ij}^s = \int_{\epsilon_{i0} > \max_k\{v_k - v_i + \epsilon_{k0}\}, \quad \epsilon_{js} > \max_k\{v_k - v_j + \epsilon_{ks}\}} dF(\epsilon_s|\epsilon_0) dF(\epsilon_0)$$

$$m_{ij}^s = \int_{\substack{\epsilon_{i0} > \max_k\{v_k - v_i + \epsilon_{k0}\}, \\ (1-\rho)\sum_{r=1}^s \Delta\epsilon_{jr} + \epsilon_{j0} > \max_k\{v_k - v_j + \sum_{r=1}^s (1-\rho)\Delta\epsilon_{kr} + \epsilon_{k0}\}}} d\Upsilon(\Delta\epsilon_s|\epsilon_{s-1}) \cdots d\Upsilon(\Delta\epsilon_2|\epsilon_1) d\Upsilon(\Delta\epsilon_1|\epsilon_0) dF(\epsilon_0)$$

Define $\kappa_j = \sum_{r=1}^s \Delta\epsilon_{jr}$. This is the cumulative change in ϵ_j relative to ϵ_i.

$$m_{ij}^s = \int_{\substack{\epsilon_{i0} > \max_k\{v_k - v_i + \epsilon_{k0}\}, \\ (1-\rho)\kappa_j + \epsilon_{j0} > \max_k\{v_k - v_j + (1-\rho)\kappa_k\}}} dF(\epsilon_0) d\Upsilon^s(\kappa|\epsilon_0)$$

Define

$$f(v_i + \epsilon_{i0} = v_j + \epsilon_{j0}) = \lim_{\rho \to 1} \frac{1}{1-\rho} \int_{\substack{v_i + \epsilon_{i0} > \max_{k \neq i,j} v_k + \epsilon_{k0} \\ v_j - v_i < \epsilon_{i0} - \epsilon_{j0} < v_j - v_i + (1-\rho)}} dF(\epsilon_0)$$

Roughly, this is a measure of the number of people that just barely prefer i to j at time 0. This is well-defined because F is differentiable.

Note that

$$\lim_{\rho \to 1} \frac{1}{1-\rho} \int_{\substack{v_i + \epsilon_{i0} > \max_{k \neq i,j} v_k + \epsilon_{k0} \\ v_j - v_i < \epsilon_{i0} - \epsilon_{j0} < v_j - v_i + (1-\rho) \\ v_k - v_i < \epsilon_{i0} - \epsilon_{k0} < v_k - v_i + A(1-\rho)}} dF(\epsilon_0) = 0$$

when $j \neq k$ for any constant A. In other words, the odds of being roughly indifferent between three places vanishes more quickly.

Take the following limit as $\rho \to 1$:

$$\lim_{\rho \to 1} \frac{m_{ij}^s}{1-\rho} = f(v_i + \epsilon_{i0} = v_j + \epsilon_{j0}) \int_{\kappa_j - \kappa_i > 0} (\kappa_j - \kappa_i) d\Upsilon^s(\kappa)$$

Note that all the $k \neq i, j$ drop out as $\rho \to 1$ because for any κ_k, the odds that it exceeds $\frac{v_j - v_k}{1-\rho} + \kappa_j$ goes to zero from the integral above. Also note that Ψ^s does not depend on ϵ_0 because of the preceding lemma.

$\Upsilon^s(\kappa)$ is the cdf of a convolution of s independent logistic distributions for each i. Since logistics are symmetrical, we can rewrite this as

$$\lim_{\rho \to 1} \frac{m_{ij}^s}{1-\rho} = f(v_i + \epsilon_{i0} = v_j + \epsilon_{j0}) \int_0^\infty \kappa d\Psi^{2s}(\kappa)$$

where Ψ^t is the convolution of t i.i.d. unidimensional logistic functions.

Finally, this implies that

$$\lim_{\rho \to 1} \frac{m_{ij}^s}{m_{ij}^1} = \frac{\int_0^\infty \kappa d\Psi^{2s}(\kappa)}{\int_0^\infty \kappa d\Psi^2(\kappa)}$$

□

A.6 Proof of Proposition 3

We intend to show that

$$\sqrt{\frac{2}{s}} \frac{\int_{-\infty}^0 \Psi^s(\kappa) d\kappa}{\int_{-\infty}^0 \Psi^2(\kappa) d\kappa}$$

is between 1 and $\sqrt{\pi/3}$ for $s \geq 2$.

We proceed in two steps. First we show that the expression of interest is increasing in s. Then we find the limit as $s \to \infty$, which is an upper bound. Since it is trivially equal to 1 for $s = 2$, that establishes the lower bound.

The expression of interest is increasing

First, we would like to show that

$$\frac{1}{\sqrt{s+1}} \int_{-\infty}^0 \Psi^{s+1}(\kappa) d\kappa \geq \frac{1}{\sqrt{s}} \int_{-\infty}^0 \Psi^s(\kappa) d\kappa$$

Note that by integrating by parts,

$$\frac{1}{\sqrt{s}} \int_{-\infty}^0 \Psi^s(\kappa) d\kappa = \frac{1}{\sqrt{s}} \int_{-\infty}^0 -\kappa \psi^s(\kappa) d\kappa = \frac{1}{2\sqrt{s}} \mathbb{E}[|X^s|]$$

where the second equality holds because the distribution is symmetric, and we define ψ^s to be the pdf, and X^s to be a random variable with distribution Ψ^s. This is useful because we will now consider the Fourier transformation of the absolute value function

$$|x| = \frac{1}{\pi} \int_{-\infty}^\infty \frac{1 - e^{-iux}}{u^2} du$$

Taking expectations,

$$\mathbb{E}[|X|] = \frac{1}{\pi} \int_{-\infty}^\infty \frac{1 - \mathbb{E}[e^{-iuX}]}{u^2} du$$

$\mathbb{E}[e^{-iuX}]$ is the characteristic function, which is known for convolutions of logistic functions. In particular, the characteristic function of a standard logistic function is:

$$\phi(u) = \frac{\pi u}{\sinh(\pi u)}$$

So the convolution of s logistic functions is:

$$\phi^s(u) = \left(\frac{\pi u}{\sinh(\pi u)}\right)^s$$

Plugging this into the Fourier transformation of the absolute value,

$$\frac{1}{\sqrt{s}}\mathbb{E}[|X^s|] = \frac{1}{\pi}\frac{1}{\sqrt{s}}\int_{-\infty}^{\infty}\frac{1-\left(\frac{\pi u}{\sinh(\pi u)}\right)^s}{u^2}du$$

A simple u-substitution:

$$\frac{1}{\sqrt{s}}\mathbb{E}[|X^s|] = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{1-\left(\frac{\pi u/\sqrt{s}}{\sinh(\pi u/\sqrt{s})}\right)^s}{u^2}du$$

And because sinh is an odd function, the integrand is an even function. So we can simplify:

$$\frac{1}{\sqrt{s}}\mathbb{E}[|X^s|] = \frac{2}{\pi}\int_{0}^{\infty}\frac{1-\left(\frac{\pi u/\sqrt{s}}{\sinh(\pi u/\sqrt{s})}\right)^s}{u^2}du$$

Now we want to prove that this is increasing in s. A sufficient condition is to prove that

$$\left(\frac{\pi u/\sqrt{s}}{\sinh(\pi u/\sqrt{s})}\right)^s$$

is decreasing in s for all u. Because we aim to prove it is true for all u, we can drop the π, and we will consider the log of the expression, which is decreasing if and only if the expression itself is decreasing.

$$f(s) = s\left(\log\left(\frac{u}{\sqrt{s}}\right) - \log\left(\frac{1}{2}\left(\exp\left(\frac{u}{\sqrt{s}}\right) + \exp\left(-\frac{u}{\sqrt{s}}\right)\right)\right)\right)$$

Even though we only care about discrete s, this equation is well-defined for continuous s, and if it is decreasing for all s that are continuous, then it is also decreasing for all discrete s.

So let us take the derivative with respect to s and show that it is negative for all u.

$$\begin{aligned} f'(s) = &\log\left(\frac{u}{\sqrt{s}}\right) - \log\left(\frac{1}{2}\left(\exp\left(\frac{u}{\sqrt{s}}\right) + \exp\left(-\frac{u}{\sqrt{s}}\right)\right)\right) \\ &+ s\left(-\frac{1}{2s} + \frac{\left(\exp\left(\frac{u}{\sqrt{s}}\right) - \exp\left(-\frac{u}{\sqrt{s}}\right)\right)}{\left(\exp\left(\frac{u}{\sqrt{s}}\right) + \exp\left(-\frac{u}{\sqrt{s}}\right)\right)}\frac{1}{2}\frac{u}{s^{3/2}}\right) \end{aligned}$$

This simplifies to:

$$\begin{aligned} f'(s) = &\log\left(\frac{u}{\sqrt{s}}\right) - \log\frac{1}{2} - \log\left(\exp\left(\frac{u}{\sqrt{s}}\right) + \exp\left(-\frac{u}{\sqrt{s}}\right)\right) \\ &- \frac{1}{2} + \frac{\exp\left(\frac{u}{\sqrt{s}}\right) - \exp\left(-\frac{u}{\sqrt{s}}\right)}{\exp\left(\frac{u}{\sqrt{s}}\right) + \exp\left(-\frac{u}{\sqrt{s}}\right)}\frac{1}{2}\frac{u}{\sqrt{s}} \end{aligned}$$

Note that

$$\log\left(\frac{u}{\sqrt{s}}\right) \leq \log 2 - 1 + \frac{1}{2}\frac{u}{\sqrt{s}}$$

because log is concave and the right-hand side is the first-order Taylor expansion of the log around 2. Also note that

$$-\log\left(\left(\exp\left(\frac{u}{\sqrt{s}}\right) + \exp\left(-\frac{u}{\sqrt{s}}\right)\right)\right) \leq -\frac{u}{\sqrt{s}}$$

Finally, note that

$$\frac{\exp\left(\frac{u}{\sqrt{s}}\right) - \exp\left(-\frac{u}{\sqrt{s}}\right)}{\exp\left(\frac{u}{\sqrt{s}}\right) + \exp\left(-\frac{u}{\sqrt{s}}\right)} \leq 1$$

Then

$$f'(s) \leq \log 2 - 1 + \frac{1}{2}\frac{u}{\sqrt{s}} - \log\frac{1}{2} - \frac{u}{\sqrt{s}} - \frac{1}{2} + \frac{1}{2}\frac{u}{\sqrt{s}} = \log 4 - \frac{3}{2} < 0$$

Therefore, $\frac{1}{\sqrt{s}}\int_{-\infty}^{0}\Psi^s(\kappa)d\kappa$ is increasing in s.

The expression of interest is bounded above by $\sqrt{\pi/3}$

Given that $\frac{1}{\sqrt{s}}\int_{-\infty}^{0}\Psi^s(\kappa)d\kappa$ is increasing in s, we next focus on its value at $s = 2$ and as $s \to \infty$.

First, consider its limiting behavior as $s \to \infty$. Here, we can apply the central limit theorem, which tells us that the sum of many logistic variables, divided by the square root of s approaches a normal variable with standard deviation equal to that of the standard logistic, known to be $\frac{\pi}{\sqrt{3}}$. Therefore, we consider

$$\int_{-\infty}^{0} \Phi\left(\frac{x\sqrt{3}}{\pi}\right) dx$$

where Φ is the standard normal distribution CDF. Integrating by parts,

$$u = \Phi\left(\frac{x\sqrt{3}}{\pi}\right), \quad v = x, \quad du = \sqrt{\frac{3}{2\pi^3}} \exp(-\frac{3x^2}{2\pi}), \quad dv = dx$$

So,

$$\begin{aligned}
\int_{-\infty}^{0} \Phi\left(\frac{x\sqrt{3}}{\pi}\right) dx &= \left[\Phi\left(\frac{x\sqrt{3}}{\pi}\right) x\right]_{-\infty}^{0} - \int_{-\infty}^{0} x\sqrt{\frac{3}{2\pi^3}} \exp(-\frac{3x^2}{2\pi}) dx \\
&= \sqrt{\frac{\pi}{6}}[e^{-x^2/2}]_{-\infty}^{0} \\
&= \sqrt{\frac{\pi}{6}}
\end{aligned}$$

We are also interested in $s = 2$, which we can calculate via integration of the logistic probability density function.

$$\begin{aligned}
\frac{1}{\sqrt{2}} \int_{-\infty}^{0} \Psi^2(\kappa) d\kappa &= \frac{1}{\sqrt{2}} \int_{-\infty}^{0} \int_{-\infty}^{\infty} \int_{-\infty}^{\kappa - x} \frac{e^{-x}}{(1+e^{-x})^2} \frac{e^{-y}}{(1+e^{-y})^2} dy dx d\kappa \\
&= \frac{1}{\sqrt{2}} \int_{-\infty}^{0} \int_{-\infty}^{\infty} \frac{e^{-x}}{(1+e^{-x})^2} \frac{1}{1+e^{-(\kappa - x)}} dx d\kappa \\
&= \frac{1}{\sqrt{2}} \int_{-\infty}^{0} \left[-\frac{e^{\kappa}(-(e^x+1)\log(e^{\kappa}+e^x) + e^{\kappa} + (e^x+1)\log(e^x+1) - 1)}{(e^{\kappa}-1)^2(e^x+1)} \right]_{-\infty}^{\infty} d\kappa \\
&= \int_{-\infty}^{0} \frac{e^{\kappa}(-\kappa + e^{\kappa} - 1)}{(e^{\kappa}-1)^2} d\kappa \\
&= \frac{1}{\sqrt{2}} \left[\frac{\kappa}{e^{\kappa}-1} + \kappa \right]_{-\infty}^{0} \\
&= \frac{1}{\sqrt{2}}
\end{aligned}$$

So the fraction

$$\lim_{s\to\infty}\sqrt{\frac{2}{s}\frac{\int_{-\infty}^{0}\Psi^s(\kappa)d\kappa}{\int_{-\infty}^{0}\Psi^2(\kappa)d\kappa}}=\frac{\sqrt{\pi/6}}{\sqrt{1/2}}=\sqrt{\frac{\pi}{3}}$$

In words, this means that the ratio we are interested will converge to $\sqrt{\pi/3}$ as s approaches ∞.

The expression of interest is bounded below by 1

We also know the ratio for $s=2$ because it is trivial:

$$\sqrt{\frac{2}{2}\frac{\int_{-\infty}^{0}\Psi^2(\kappa)d\kappa}{\int_{-\infty}^{0}\Psi^2(\kappa)d\kappa}}=1$$

Completing the proof

And since $\frac{1}{\sqrt{s}}\int_{-\infty}^{0}\Psi^s(\kappa)d\kappa$ is increasing in s it follows that

$$1\leq\sqrt{\frac{2}{s}\frac{\int_{-\infty}^{0}\Psi^s(\kappa)d\kappa}{\int_{-\infty}^{0}\Psi^2(\kappa)d\kappa}}\leq\sqrt{\frac{\pi}{3}}$$

for all $s\geq 2$.

Finally, we substitute in the expression from Lemma 1:

$$1\leq\frac{1}{\sqrt{s}}\frac{m_{ij}^s}{m_{ij}^1}\leq\sqrt{\frac{\pi}{3}}$$

□

A.7 Proof of Proposition 4

When $\Delta\to\infty$, the migration rate is given by:

$$\frac{m_{i\to j}}{p_i}=\exp(v_j-v_i-\delta_{ij})$$

So

$$\frac{\partial\log m_{i\to k}}{\partial v_k}-\frac{\partial\log p_i}{v_k}=1$$

and

$$\frac{\partial\log m_{k\to i}}{\partial v_k}-\frac{\partial\log p_k}{v_k}=-1$$

In any steady-state, it must be the case that total inmigration and total outmigration are equal, which means that

$$\sum_{i \neq k} \left(\frac{\partial m_{i \to k}}{\partial v_k} - \frac{\partial m_{k \to i}}{\partial v_k} \right) = 0$$

For all $k \neq j$, the change in $\partial \log p_j / \partial v_k$ must be the same. This is because the migration rate between i and j does not change with v_k when $\Delta \to \infty$. So if their relative populations change, then the migration levels will change and we will not be in steady-state.

Substituting in,

$$\left(\sum_{i \neq k} m_{i \to k} \right) \left(\frac{\partial p_i}{\partial v_k} + 1 - \left(\frac{\partial p_k}{\partial v_k} - 1 \right) \right) = 0$$

where we can pull out the $\partial p_i / \partial v_k$ since it is the same for all i. This simplifies to:

$$\frac{\partial \log p_k}{\partial v_k} - \frac{\partial \log p_j}{\partial v_k} = 2$$

The total population is fixed:

$$p_k \frac{\partial \log p_k}{\partial v_k} + (1 - p_k) \frac{\partial \log p_j}{\partial v_k} = 0$$

So we need to solve for the population elasticities with two equations and two unknowns. The solution is:

$$\frac{\partial \log p_k}{\partial v_k} = 2 - 2p_k$$
$$\frac{\partial \log p_j}{\partial v_k} = -2p_k$$

□

B Appendix Tables and Figures

B.1 Distribution of the Square Root Fact

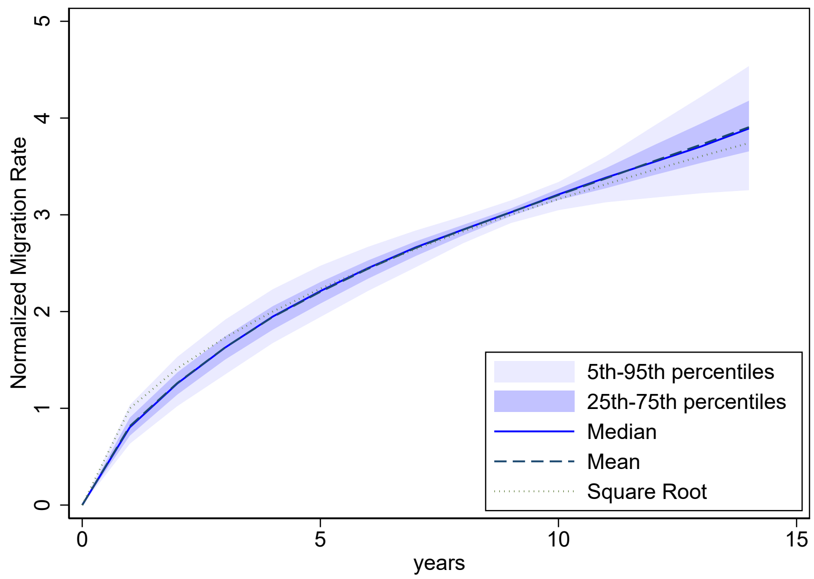
In this subsection, we examine how the square root fact is distributed across state pairs and age cohorts to see if it holds generally or is specific to our aggregation.

To create Figure A1a, we calculate $m^t_{i\to j}$, the migration rate from state i to state j at time t, for all state pairs. Because migration volumes vary significantly between different pairs (e.g., migration between California and Texas is much higher than between Wyoming and Vermont), we apply a normalization to ensure comparability. Specifically, we use the factor

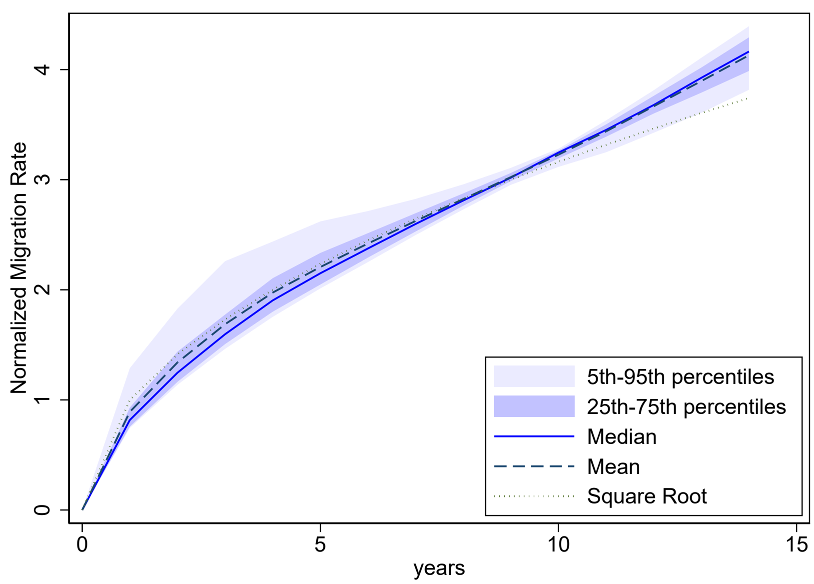
$$c_{ij} = \frac{\sum_{x=1}^{14} \sqrt{x}}{\sum_{t=1}^{14} m^t_{i\to j}},$$

which scales the migration rates relative to the square root benchmark. We then plot the 5th, 25th, 50th (median), 75th, and 95th percentiles, along with the mean, of the normalized migration rates $c_{ij} m^t_{i\to j}$ across all state pairs. These values are weighted by the total migration volume $\sum_{t=1}^{14} m^t_{i\to j}$ for each pair. The results indicate that the square root fact holds consistently across state pairs, with relatively little deviation.

A similar analysis is presented in Figure A1b, where we examine the square root fact across different age cohorts rather than state pairs. Again, the distribution closely aligns with the square root relationship, though there is slightly more variation compared to the state-pair analysis. Notably, we observe a “bulge” above the square root line at shorter horizons. This deviation is driven by younger cohorts.



(a) Square Root Fact across state-pairs. For each state-pair, we calculate the share of people that move from state i to state j over t years. We then normalize these shares such that the sum of the shares over t equals $\sum_{x=1}^{14} \sqrt{s}$. This figure shows the t-by-t distribution over the (i, j) pairs, weighted by the number of migrants from i to j.

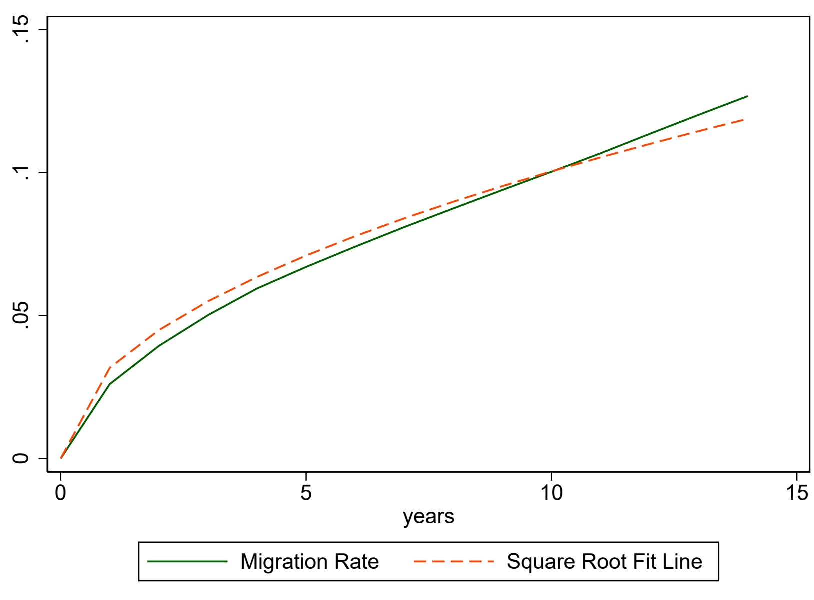


(b) Square Root Fact across age cohorts. For each age cohort, we calculate the share of people that move from state i to state j over t years. We then normalize these shares such that the sum of the shares over t equals $\sum_{x=1}^{14} \sqrt{s}$. This figure shows the t-by-t distribution over the age cohorts, weighted by the size of the cohort.

B.2 Age and the Square Root Fact

A reader might worry that migration in the GCCP is not measured well for young people. For example, college students may not be measured accurately. In the GCCP, the data provider includes a proxy for age, which we use to exclude people that are young. As you can see in Figure A2a, younger people are much more likely to move, and so it is a reasonable concern that the square root fact may be specific to young people. However, in Figure A2b, we can exclude people below the age of 45, and the square root fact still holds.

(a) Moving probability by age in the GCCP



(b) Square root fact for people over 45 in the GCCP

Figure A2: Robustness to considering whether age explains the square root fact

B.3 Long-run population elasticities in the moving cost model

In this appendix, we compare the elasticities in the calibrated moving cost model to the approximation from Proposition 4, which stated that the long-run population elasticities were proportional to those of a standard static logit model.

Calculating these elasticities for the moving cost model is not straightforward analytically, but easy to do numerically. For each state, we change the utility by a small amount, calculate the new migration matrix from equation (3), and simulate 500 periods to see how much population in each state changes.

The overall correlation between the long-run elasticities of the moving cost model and the static logit is 0.99996. Splitting it between same-state and cross-state elasticities, the correlation is 0.948 and 0.9994, respectively. Hence, it is a reasonable approximation to say that the moving cost model approaches a static logit in the long-run.

We show a plot of the cross-state elasticities in Figure A3. The relationship is quite strong.

Recall that, in contrast, the SPACE model does not approach a static logit model, but has a much richer set of cross-elasticities, given by the migration shares.

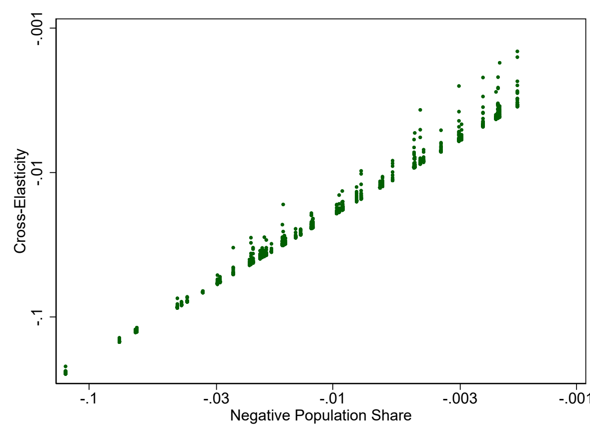


Figure A3: Population cross-elasticities in simulations of the moving cost model versus the theoretical approximation (in a static logit, the cross-elasticity is proportional to the population share)

C Non-Steady-State Migration

In our calibration, we can still solve for migration even if the model is not in steady-state. In particular, in a two-region model where the ϵ's are independent across space and the correlation across time is governed by $\tilde{\rho}$, migration is from i to j is defined as:

$$m_{i\to j} = P(\epsilon_{i1} + v_{i1} > \epsilon_{j1} + v_{j1}, \epsilon_{i2} + v_{i2} < \epsilon_{j2} + v_{j2})$$

where

$$F(\epsilon_{i1}, \epsilon_{j1}, \epsilon_{i2}, \epsilon_{j2}) = \exp\left(-\left(e^{-\frac{\epsilon_{i1}}{1-\tilde{\rho}}} + e^{-\frac{\epsilon_{i2}}{1-\tilde{\rho}}}\right)^{1-\tilde{\rho}} - \left(e^{-\frac{\epsilon_{j1}}{1-\tilde{\rho}}} + e^{-\frac{\epsilon_{j2}}{1-\tilde{\rho}}}\right)^{1-\tilde{\rho}}\right)$$

Let us define $v_1 = v_{i1} - v_{j1}$ and $v_2 = v_{i2} - v_{j2}$. Without loss of generality, assume $v_2 > v_1$. This implies that $m_{j\to i} > m_{i\to j}$ Then the probability is given by

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\partial^2 F}{\partial \epsilon_{i1}\epsilon_{j2}}(\epsilon_{i1}, \epsilon_{i1} - v_1, \epsilon_{j2} + v_2, \epsilon_{j2}) d\epsilon_{i1} d\epsilon_{j2}$$

Written out,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(-\left(e^{-\frac{\epsilon_{i1}}{1-\tilde{\rho}}} + e^{-\frac{\epsilon_{j2}+v_2}{1-\tilde{\rho}}}\right)^{1-\tilde{\rho}} - \left(e^{-\frac{\epsilon_{i1}-v_1}{1-\tilde{\rho}}} + e^{-\frac{\epsilon_{j2}}{1-\tilde{\rho}}}\right)^{1-\tilde{\rho}}\right)$$
$$\times \left(e^{-\frac{\epsilon_{i1}}{1-\tilde{\rho}}} + e^{-\frac{\epsilon_{j2}+v_2}{1-\tilde{\rho}}}\right)^{1-\tilde{\rho}} e^{-\frac{\epsilon_{i1}}{1-\tilde{\rho}}} \left(e^{-\frac{\epsilon_{i1}-v_1}{1-\tilde{\rho}}} + e^{-\frac{\epsilon_{j2}}{1-\tilde{\rho}}}\right)^{1-\tilde{\rho}} e^{-\frac{\epsilon_{j1}}{1-\tilde{\rho}}} d\epsilon_{i1} d\epsilon_{j2}$$

If we substitute

$$\begin{aligned}
u &= -(e^{-\frac{\epsilon_{i1}}{1-\tilde{\rho}}} + e^{-\frac{\epsilon_{j2}+v_2}{1-\tilde{\rho}}})^{1-\tilde{\rho}} \\
w &= -(e^{-\frac{\epsilon_{i1}-v_1}{1-\tilde{\rho}}} + e^{-\frac{\epsilon_{j2}}{1-\tilde{\rho}}})^{1-\tilde{\rho}} \\
dudw &= \left(1 - \exp\left(\frac{v_1 - v_2}{1-\tilde{\rho}}\right)\right)\left(e^{-\frac{\epsilon_{i1}}{1-\tilde{\rho}}} + e^{-\frac{\epsilon_{j2}+v_2}{1-\tilde{\rho}}}\right)^{1-\tilde{\rho}} e^{-\frac{\epsilon_{i1}}{1-\tilde{\rho}}} \left(e^{-\frac{\epsilon_{i1}-v_1}{1-\tilde{\rho}}} + e^{-\frac{\epsilon_{j2}}{1-\tilde{\rho}}}\right)^{1-\tilde{\rho}} e^{-\frac{\epsilon_{j1}}{1-\tilde{\rho}}} d\epsilon_{i1} d\epsilon_{j2}
\end{aligned}$$

the integral becomes

$$\frac{1}{1 - \exp\left(\frac{v_1 - v_2}{1-\tilde{\rho}}\right)} \int_{-\infty}^{0}\int_{\exp(-v_1)w}^{\exp(-v_2)w} \exp(u + w) du dw$$

Taking the interior integral,

$$\frac{1}{1 - \exp(\frac{v_1 - v_2}{1-\tilde{\rho}})} \int_{-\infty}^{0} \exp((e^{-v_2} + 1)w) - \exp((e^{-v_1} + 1)w) dw$$

Taking the second integral,

$$\frac{1}{1-\exp(\frac{v_1-v_2}{1-\tilde{\rho}})}\left(\frac{1}{1+e^{-v_2}}-\frac{1}{1+e^{-v_1}}\right)$$

This represents the share of $\tilde{w}_{ij}+\tilde{w}_{ji}$ people that move from i to j. The expression for the other direction can be found by switching v_1 for $-v_1$ and v_2 for $-v_2$.

We are interested in the amount of migration there would have been had we been in a steady-state at v_2, since it is steady-state migration that pins down the parameters of $\tilde{w}_{ij}$ and, more importantly, the cross-elasticities of population. This is given by

$$m_{ij}^{ss}=(1-\tilde{\rho})\frac{e^{v_2}}{(1+e^{v_2})^2}$$

Given

$$\begin{aligned} m_{ij} &= \frac{1}{1-\exp(\frac{v_1-v_2}{1-\tilde{\rho}})}\left(\frac{1}{1+e^{-v_2}}-\frac{1}{1+e^{-v_1}}\right) \\ m_{ji} &= \frac{1}{1-\exp(\frac{v_2-v_1}{1-\tilde{\rho}})}\left(\frac{1}{1+e^{-v_1}}-\frac{1}{1+e^{-v_2}}\right) \end{aligned}$$

which we observe in the data, and a calibrated $1-\tilde{\rho}$, we can solve for v_2, and therefore find m_{ij}^{ss}.

Note that $\frac{m_{ij}}{m_{ji}}=\exp\left(\frac{v_2-v_1}{1-\tilde{\rho}}\right)$. So $e^{-v_1}=e^{-v_2}(m_{ij}/m_{ji})^{1-\tilde{\rho}}$. Then

$$m_{ij}=\frac{1}{1-m_{ji}/m_{ij}}\left(\frac{1}{1+e^{-v_2}}-\frac{1}{1+e^{-v_2}(m_{ij}/m_{ji})^{1-\tilde{\rho}}}\right)$$

Simplifying,

$$m_{ij}-m_{ji}=\frac{e^{-v_2}((m_{ij}/m_{ji})^{1-\tilde{\rho}}-1)}{(1+e^{-v_2})(1+e^{-v_2}(m_{ij}/m_{ji})^{1-\tilde{\rho}})}$$

At this point, we could solve a quadratic equation for e^{-v_2} and plug that into m_{ij}^{ss}.[39] However, there is not much economic intuition in that solution, so here we also consider

[39] In particular, $e^{-v_2}=\frac{-b-\sqrt{b^2-4ac}}{2a}$ where

$$\begin{aligned} a &= (m_{ij}-m_{ji})(m_{ij}/m_{ji})^{1-\tilde{\rho}} \\ b &= (m_{ij}-m_{ji})+(m_{ij}-m_{ji})(m_{ij}/m_{ji})^{1-\tilde{\rho}}-(m_{ij}/m_{ji})^{1-\tilde{\rho}}+1 \\ c &= m_{ij}-m_{ji} \end{aligned}$$

and $m_{ij}^{ss}=(1-\tilde{\rho})\frac{e^{-v_2}}{(1+e^{-v_2})^2}$. Note that the other root of the quadratic equation gives you e^{-v_1}, which without loss of generality, we had assumed to be larger.

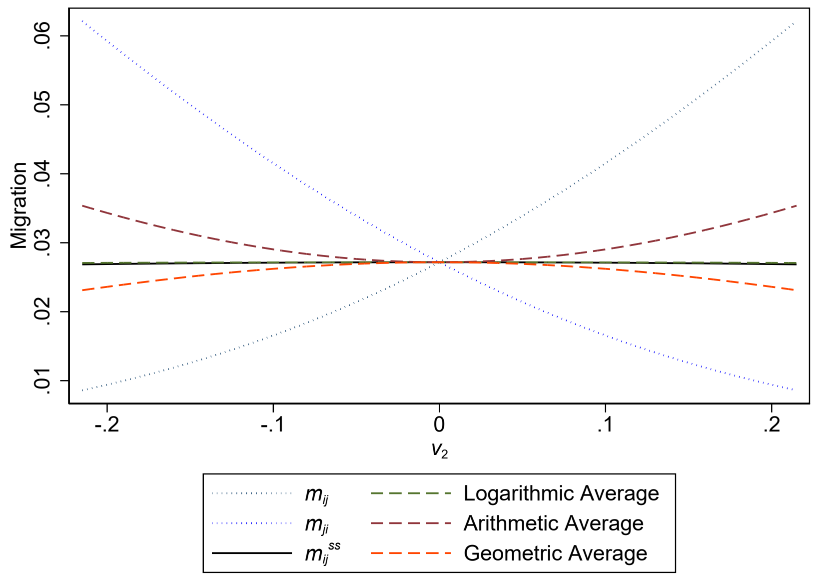


Figure A4: Bilateral migration outside of steady-state

an approximation when $\tilde{\rho} \approx 1$. In particular, we can rearrange the equation to be

$$\frac{m_{ij} - m_{ji}}{\left(\frac{(m_{ji}/m_{ij})^{1-\tilde{\rho}}-1}{1-\tilde{\rho}}\right)} = m_{ij}^{ss} \frac{1 + e^{-v_2}}{1 + e^{-v_2}(m_{ji}/m_{ij})^{1-\tilde{\rho}}}$$

If we take the limit of both sides as $\tilde{\rho} \to 1$, then

$$\frac{m_{ij} - m_{ji}}{\log(m_{ij}/m_{ji})} = \lim_{\tilde{\rho}\to 1} m_{ij}^{ss}$$

So when $\tilde{\rho} \approx 1$, the logarithmic mean of the migrations is a good approximation of the steady-state migration.

In Figure A4, we plot m_{ij}^{ss} along with m_{ij} and m_{ji}, and several candidate approximations. To create this figure, we calculate m_{ij}, m_{ji} and m_{ij}^{ss} using the formulae above, assuming that $v_1 = 0$ and that v_2 ranges from -0.25 to 0.25. This implies that the period 1 populations are 0.5 in both i and j, and each population ranges from about 0.45 to 0.55 in period 2. We then plot the logarithmic, the arithmetic, and the geometric average of m_{ij} and m_{ji} for each value of v_2, to be able to compare how good of an approximation each one is. The logarithmic approximation is visually indistinguishable from the true m_{ij}^{ss}. The other two are good approximations if $v_2 \approx 0$, but there are sizable gaps for larger v_2.